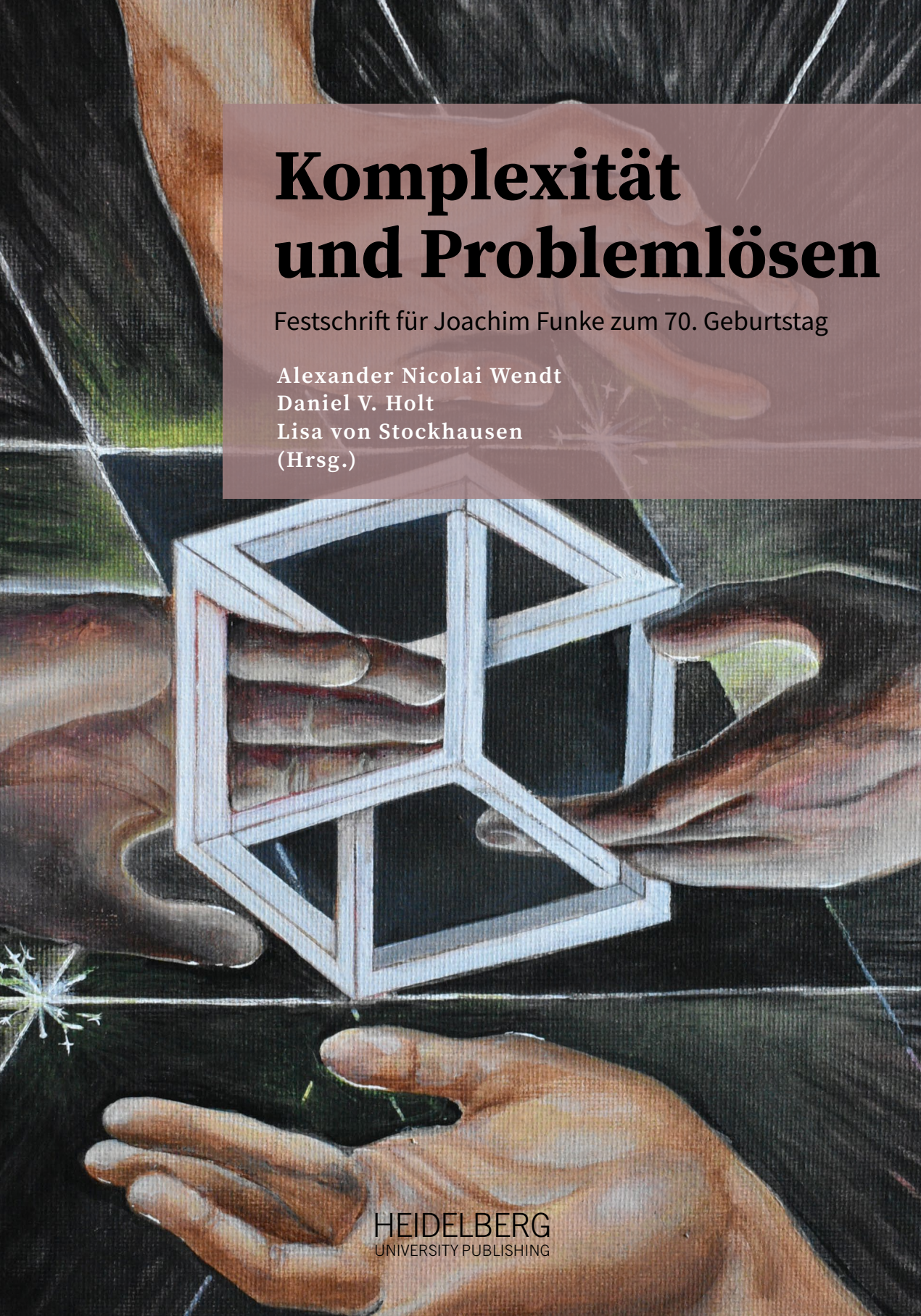




Komplexität und Problemlösen

Festschrift für Joachim Funke zum 70. Geburtstag

Alexander Nicolai Wendt
Daniel V. Holt
Lisa von Stockhausen
(Hrsg.)

HEIDELBERG
UNIVERSITY PUBLISHING

Komplexität und Problemlösen.
Festschrift für Joachim Funke zum 70. Geburtstag

Komplexität und Problemlösen

**Festschrift für Joachim Funke zum
70. Geburtstag**

**Alexander Nicolai Wendt, Daniel V. Holt &
Lisa von Stockhausen (Hrsg.)**

Alexander Nicolai Wendt (SFU Wien):  <https://orcid.org/0000-0002-2885-3609>

Daniel V. Holt (Universität Heidelberg):  <https://orcid.org/0000-0003-3616-638X>

Lisa von Stockhausen (Universität Duisburg-Essen):  <https://orcid.org/0000-0003-4774-2872>

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet unter <https://dnb.dnb.de> abrufbar.



Dieses Werk ist unter der Creative-Commons-Lizenz CC BY 4.0 veröffentlicht. Die Umschlaggestaltung unterliegt der Creative-Commons-Lizenz CC BY-ND 4.0.

Publiziert bei Heidelberg University Publishing (heiUP), 2025

Universität Heidelberg / Universitätsbibliothek

Heidelberg University Publishing (heiUP)

Grabengasse 1, 69117 Heidelberg

<https://heiup.uni-heidelberg.de>

E-Mail: ub@ub.uni-heidelberg.de

Die Online-Version dieser Publikation ist auf den Verlagswebseiten von Heidelberg University Publishing <https://heiup.uni-heidelberg.de> dauerhaft frei verfügbar (Open Access).

urn: [urn:nbn:de:bsz:16-heiup-book-1067-3](https://nbn-resolving.org/urn:nbn:de:bsz:16-heiup-book-1067-3)

doi: <https://doi.org/10.17885/heiup.1067>

Text © 2025, das Copyright der Texte liegt bei den jeweiligen Verfasser:innen.

Umschlagabbildung: © Mariangel Beatriz Mendoza de Wendt

ISBN 978-3-96822-172-4 (Hardcover)

ISBN 978-3-96822-171-7 (PDF)

Inhaltsverzeichnis

1	Joachim Funke als Denkpsychologe. Editorial ALEXANDER NICOLAI WENDT, DANIEL V. HOLT UND LISA VON STOCKHAUSEN	1
<i>Teil I: Biografisches</i>		
2	Joachim Funke: 15 Jahre Wegbegleiter JÜRGEN BREDEKAMP	15
3	Joachim Funke verarbeitet Informationen ANNEMARIE FRITZ, WALTER HUSSY UND DAVID TOBINSKI	21
4	Zeit mit Jofu: Immer dynamisch, gelegentlich komplex, nie problematisch. Reminiszenzen zweier, die es wissen sollten JENS F. BECKMANN UND AXEL BUCHNER	35
5	Ein „Funkenschlag“ für die Geschichte der Psychologie SUSANNE GUSKI-LEINWAND	47
<i>Teil II: Essays</i>		
6	Drei Jahrzehnte Gummibärenforschung: Entwicklungstrends und neue Ergebnisse JOCHEN MUSCH UND EDGAR ERDFELDER	57
7	The Role of Credulity in Failed Problem Solving ROBERT J. STERNBERG	79

INHALTSVERZEICHNIS

8	The Nature of Research on Problem Solving BENŐ CSAPÓ	101
9	Die vier Wellen der psychometrischen Forschung zu komplexen Problemen MATTHIAS STADLER	121
10	Die Problemlösung ist noch lange nicht die Lösung des Problems ANNETTE KÄMMERER	131
11	Die Universität ist kein Wirtschaftsunternehmen! BEN ROEG UND PETA PANTA	137
<i>Teil III: Originalia</i>		
12	Problemlösen und implizites Lernen HILDE HAIDER	149
13	Implicit Learning Across Isoproblems in the Tower of London Puzzle is Impaired in Patients with Parkinson's Disease AMORY H. DANEK	157
14	Metacognition and Climate Change: Why Dealing with Politicized Science Requires Self-Insight HELEN FISCHER AND DOROTHEE AMELUNG	181
15	Pragmatics—A Rich Domain for the Study of Meta-Reasoning KLAUS FIEDLER AND SHIRA ELQAYAM	191
16	Can Self-Explanations and Thinking Aloud Promote Performance at the Tower of Hanoi Problem? CHRISTINE BLECH	217
17	Complex Problem Solving and Dynamic Decision Making: What Is the Difference? CLEOTILDE GONZALEZ	233
18	Psychologische Theorien als finite Automaten WOLFGANG SCHOPPEK	249

19	Human Eye Movements as Indicators of Complexity	
	MIRIAM SPERING	269
	Verzeichnis der Autorinnen und Autoren	287

Joachim Funke als Denkpsychologe. Editorial

ALEXANDER NICOLAI WENDT^{1,2}, DANIEL V. HOLT¹ UND LISA VON
STOCKHAUSEN³

¹Ruprecht-Karls-Universität Heidelberg; ²Sigmund Freud PrivatUniversität Wien, Österreich;

³Universität Duisburg-Essen

So selbstverständlich es für die Psychologie des beginnenden 20. Jahrhunderts gewesen ist, von der Berliner Schule der Gestaltpsychologie, der Zweiten Leipziger Schule der Ganzheitspsychologie oder der Würzburger Schule der Denkpsychologie zu sprechen – um nur die bekanntesten Beispiele zu nennen –, so ungewöhnlich ist die akademische Schulbildung in der Psychologie des frühen 21. Jahrhunderts geworden. Dafür gibt es vielfältige Gründe, die dem Wandel der Wissenschaftskultur im Laufe des letzten Jahrhunderts entsprechen: Karrieren sind, so wie das gesamte Leben, mobiler geworden. Selten beschränkt sich ein Forschungsnetzwerk auf eine geografische Region. Im Gegenteil gestattet die Infrastruktur der Nachkriegszeit und spätestens des Computerzeitalters, die Ressourcen und Perspektiven diverser Standorte zu verbinden. Auch ist die Identifikation mit Forschungstraditionen aus der Mode gekommen.

Die in der Regel vorherrschende wissenschaftstheoretische Grundüberzeugung ist es, dass Theorien kritisch überprüft und möglichst sparsam aufgestellt sein sollten. Forschungsprogramme beschränken sich selten auf das Lehrgebäude einer

Schule, sondern entwickeln ihre Modelle dynamisch und orientieren sich stärker an Problemen und der Entwicklung von Methoden zu ihrer Bewältigung als an der Arbeit an einem Gesamtsystem. Nicht zuletzt ist der Anspruch der historischen Herleitung der eigenen Forschung hinter die Absicherung naturwissenschaftlicher Exaktheit zurückgetreten. Aus diesen und weiteren Gründen ist es heutzutage ungewöhnlich, eine psychologische Forschungsbiografie durch Schulenzugehörigkeit zu charakterisieren.

Obwohl im Fach Psychologie die Zeichen der Zeit nicht zugunsten der wissenschaftsgeschichtlichen Einordnung stehen mögen, so gibt ein runder Geburtstag doch Anlass, zurückzublicken und ein Lebenswerk zu würdigen. Wie kann das aber gelingen? Wie ermisst sich die Bedeutung und die Geltung einer wissenschaftlichen Leistung? Schwerlich ist in diesen Fragen eine historische Rückschau zu vermeiden, da doch selbst dann, wenn sich in der Psychologie alles um die inkrementelle Bewährung von Theorien an empirischen Schlüsselexperimenten drehen sollte, dieser Erkenntnisfortschritt und Wissenszuwachs nur relativ ist, an keiner absoluten Schwelle anlangen kann. Wissenschaft ist ein nicht abreißender Strom neuer Einsichten und Probleme, sodass wir nur begreifen, ob wir vorangekommen sind, indem wir uns auf die bereits zurückgelegte Strecke besinnen.

Wissenschaftsgeschichtsschreibung ist nicht nur als Personen- oder Institutionengeschichte, sondern insbesondere auch als Ideengeschichte möglich (siehe Pongratz, 1967). Ein Jubiläum gibt Anlass, die Entwicklungen eines Themenbereichs zu betrachten. In Joachim Funks Fall liegt es nahe, in dieser Frage auf das Feld der psychologischen Problemlösungsforschung zu blicken. Mit dem theoretischen Schritt in Richtung der Komplexität und erheblicher methodologischer Kreativität hat Joachim Funke über mehrere Jahrzehnte einen substanziellen Beitrag geleistet, um sowohl empirische Einsichten zu ermöglichen als auch die Frage nach dem Sinn des Problemlösens selbst besser zu verstehen. Die zahlreichen Publikationen Funks haben jedoch auch noch eine andere Signatur: Neben das Wort *Problemlösen* tritt das Wort *Denken*. Das ist zunächst nicht überraschend, da die Begriffe inhaltlich eng zusammengehören. Dennoch ist zu bestimmen, inwieweit Joachim Funks Werk als eine genuine Form der *Psychologie des Denkens*

verstanden werden kann, die über aktuelle Trends und Paradigmen der kognitiven Psychologie hinausweist.

Was der Blick auf die Probleme, denen sich Wissenschaftlerinnen und Wissenschaftler widmen, zu übersehen droht, ist indes, wie jemand auf ein Thema gekommen ist. Die Ideengeschichte setzt zumeist bei der Idee an, fragt aber nicht danach, was die Ideen selbst relevant macht. Hierfür ist Personen- und insbesondere Institutionengeschichte besser geeignet als eine reine Ideengeschichte. Um zu verstehen, was Joachim Funke darauf gebracht hat, komplexe Probleme zu erforschen, ist der Blick auf die Traditionen wertvoll, mit denen er in Beziehung gestanden hat. Auch wenn der Schulenbegriff im Diskurs der Disziplin keine Rolle mehr zu spielen scheint, existiert letztlich doch ein Netzwerk der Forschungstraditionen wie ein soziales Wurzelwerk der Gedanken, ein Rhizom der Wissenschaftsbio- grafien zwischen Forscherinnen- und Forschergenerationen. So tritt neben die ideengeschichtliche Hypothese eine institutionsgeschichtliche: Joachim Funke ist ein *Denkpsychologe*.

Ursprünge der Denkpsychologie

Oftmals wird der Begriff *Würzburger Schule* mit der Denkpsychologie gleich- gesetzt, wobei besonders Oswald Külpe, dem Gründer des Würzburger Instituts für Psychologie, eine hervorgehobene Rolle zufällt. Weitere prägende Mitglieder dieser geistigen Bewegung waren Narziß Ach, Karl Bühler, Johannes Lindworsky, Karl Marbe und Otto Selz. Auch wenn Würzburg als Ort eine zentrale Bedeutung für die Geschichte der denkpsychologischen Forschungsgemeinschaft zukommt, entwickelte sie sich auch andernorts. Als ein prominentes Beispiel sei Alfred Binet in Frankreich genannt, ein Pionier der Messung kognitiver Fähigkeiten im Bil- dungskontext, der sich bereits Ende des 19. Jahrhunderts introspektiver Methoden bediente. Die Forschungstradition der Denkpsychologie umfasst also mehr als eine regionale Gruppierung oder eine wissenschaftliche Ahnenfolge. Im Kern stehen ein methodologischer Ansatz, der durch den Terminus *systematische experimentelle Selbstbeobachtung* ausgedrückt werden kann, sowie ein psychisches Phänomen,

nämlich der sogenannte *anschauungslose Gedanke*. Damit ist gemeint, dass nicht alle Denkerfahrungen wahrnehmungsnahe (z. B. visuelle oder akustische) Vorstellungen sind. Ein typisches Beispiel ist das Regelbewusstsein: Es ordnet das Denken, auch wenn ihm kein Bild oder Wort entspricht. Dieses Untersuchungsfeld ist von großer historischer Bedeutung, weil es der im 19. Jahrhundert einflussreichsten psychologischen Weltanschauung widerspricht, also der Assoziationspsychologie. Anders gesagt: Mittels systematischer introspektiver Verfahren versucht die Denkpsychologie eine Klasse von Erfahrungen und mentalen Prozessen zu bestimmen, die nicht auf grundlegende Elemente reduziert werden können, sondern holistisch sind. In gewisser Hinsicht handelt es sich um einen Ansatz zur Erforschung der Psychologie höherer kognitiver Funktionen.

Hier ließe sich einwenden, dass höhere kognitive Funktionen ein universales Phänomen sind, das vielfach wiederentdeckt wurde und es deswegen keiner Tradition bedürfe, um sie zu bestimmen. Was daran stimmt, ist ohne Zweifel, dass die Selbstreflexion ein menschliches Grundvermögen ist und es deswegen jederzeit möglich ist, den Gegenstandsbereich der Denkpsychologie immer wieder neu zu erschließen. Was allerdings nicht unterschätzt werden sollte, ist die Bedeutung theoretischer und methodischer Errungenschaften, welche die Denkpsychologie von der naiven Alltags-Introspektion abgrenzen und in den Rang einer wissenschaftlichen Disziplin erheben. Ebenso wie das Prinzip der Lichtbrechung jederzeit wiederentdeckt werden kann, sobald ein Mensch die plötzliche Veränderung des visuellen Eindrucks bemerkt, wenn ein Objekt ins Wasser eintaucht, ist es doch historisch unstrittig, dass die Einsichten und Methoden der frühen Naturwissenschaften prometheische Leistungen gewesen sind. Im Hinblick auf die Lichtbrechung ist beispielsweise Abu Ali al-Hasan ibn al-Haitham (Alhazen, 945–1040) als Pionier der experimentell-induktiv arbeitenden Optik einer der Riesen, auf dessen Schultern wir stehen. Dasselbe gilt auch für die psychologische Theoriebildung. Deswegen ist die unabhängige konvergente Wiederentdeckung¹

¹ Der Begriff *unabhängige konvergente Wiederentdeckung* wird hier neu geprägt und meint, dass derselbe Gedanke (resp. dieselbe Theorie) zu einem späteren Zeitpunkt erneut identisch gedacht

ehemaliger Theoriegebäude keine Trivialität und Traditionen bestimmen vielfach – bemerkt oder unbemerkt – die Psychologiegeschichte.

Joachim Funke im Kontext der Denkpsychologie

Die zentrale Schaffensphase der Denkpsychologie fiel in die ersten beiden Jahrzehnte des 20. Jahrhunderts. Es muss deswegen konstruiert erscheinen, Joachim Funke in diese Tradition zu stellen. Und doch gibt es wissenschaftsgeschichtliche Gründe, in der Psychologie des komplexen Problemlösens ein Erbe der ursprünglichen Denkpsychologie zu sehen: Neben dem familiären Stammbaum steht der akademische, ein Privileg, das mit dem Doktorgrad erworben wird. In Joachim Funks Fall sind hier zwei Namen zu nennen: Seine Promotion wurde von Jürgen Bredenkamp und Walter Hussy betreut. Beide sind in gewissem Sinne mit der denkpsychologischen Tradition verbunden.

Jürgen Bredenkamp ist schwerlich als Denkpsychologe zu bezeichnen. Dennoch gibt es eine direkte personelle Beziehung. Bredenkamp wurde 1964 in Heidelberg bei Carl Friedrich Graumann als Assistent eingestellt, um dort 1971 mit einer methodologischen Arbeit habilitiert zu werden. In der Festschrift für Graumanns 80. Geburtstag spricht Funke deswegen von „Jürgen Bredenkamp, der ein Graumann-Schüler und Heidelberger Alumnus ist“ (Funke, 2003a, S. 122). Graumann nun verstand sich selbst als Denkpsychologe, wurde er doch 1952 mit einer denkpsychologischen Arbeit über das *Einfallserleben* bei Maria Krudewig promoviert, die ihrerseits Schülerin Johannes Lindworskys und damit eines (zumindest mittelbaren, da Kölner) Mitglieds der Würzburger Schule gewesen ist.

wird, ohne dass ein historischer Bezug zum vorigen Mal bestünde. Der Begriff ist an das evolutionistische Konzept der *konvergenten Evolution* angelehnt, bei der an Organismen genotypisch oder phänotypisch identische biologische Strukturen in geografisch verbindungslosen Ahnenreihen entstehen, weil analoge Umwelthanforderungen bestehen. Auch wenn konvergente Evolution in einigen Fällen beobachtet worden ist, handelt es sich um ein ungewöhnliches Phänomen und es scheint uns, dass auch für theoretische Einfälle gilt, dass ihr ko-inzidentelles Auftreten unwahrscheinlich ist.

Verfolgen wir Bredenkamps akademische Abstammung weiter, tritt noch eine erstaunliche historische Verbindung zutage: Bredenkamp wurde von Peter Hofstätter promoviert, der wiederum – wie übrigens auch Karl Popper – seinen Doktorgrad unter Karl Bühler erlangte, einem unmittelbaren Schüler Oswald Külpes. Die Konvergenzlinien mögen im Sand der Zeit verweht sein, doch im geistesgeschichtlichen Untergrund ist Kontinuität bestehen geblieben.

Der Bezug zur Denkpsychologie ist im Falle Walter Hussys offensichtlicher, etwa dort, wo Hussy im Vorwort seines 1984 veröffentlichten zweibändigen Lehrbuchs der Denkpsychologie ausdrücklich „Herrn Dipl. Psych. J. Funke ...für seine unermüdliche Diskussionsbereitschaft und zahlreichen Hinweise“ (Hussy, 1984, S. 11) dankt. Allerdings ist Hussys Denkpsychologie auf eine andere Weise in die Tradition integriert. Die Würzburger Schule wird bei Hussy historisch gewürdigt, wichtiger aber ist für seine Forschung, wie bereits aus dem Titel seiner Dissertation von 1973 hervorgeht, die *Informationsverarbeitung als reduzierender und generierender Prozess* (Hussy, 1973). Es handelt sich um kognitivistische Terminologie, deren Kontext an den US-amerikanischen Quellen in Hussys Texten dieser Zeit sichtbar wird. Der neuralgische Bezugspunkt ist die Forschung von Allen Newell und Herbert Simon und interessanterweise hat auch diese Forschung einen expliziten Bezug zur Würzburger Denkpsychologie vorzuweisen, wie eine Rezension Newells verrät:

Here is a translation of a book written in the early '40's by a Dutch psychologist working within a framework created by a German psychologist, Otto Selz, whose work was published in the early '20's. Can it be relevant to living science, anno 1967? The answer is yes, and thereby hangs partly a minor tale in the history of psychology but mostly a tale of the viability of data when the time is ripe. (Newell, 1967, S. 803)

Die Rede ist von Adriaan de Groot's 1946 verfasster Doktorarbeit unter dem Titel *Het Denken van den Schaker* (de Groot, 1946). De Groot hatte die Gelegenheit, bei Otto Selz, der vor seiner Deportation nach Amsterdam emigrierte, die Me-

thodologie und Theorie der Denkpsychologie kennenzulernen, und wandte sie daraufhin auf die Domäne des Schachspiels an. Diese Forschung war Newell und Simon vertraut und inspirierte ihre eigene Forschung, wie ihre empirischen Untersuchungen zum Schachspiel in *Human Problem Solving* (Newell & Simon, 1972) beweisen. Die Problemlösungsforschung trägt auf diesem, aber auch auf weiteren Wegen das Erbe der Denkpsychologie (siehe Frijda & de Groot, 1981). In Walter Hussys Auseinandersetzung mit der Informationsverarbeitung, die Joachim Funke schon Jahre zuvor aktiv forschend begleitet hat – wie sich anhand eines Trierer Psychologischen Berichts mit dem Titel *KOGNI: Ein Computerprogramm für die Berechnung verschiedener Maße zur Erfassung kognitiver Komplexität* (Funke & Hussy, 1977) nachvollziehen lässt –, setzt sich letztlich ebenso eine denkpsychologische Tradition fort.

Die Zukunft der Denkpsychologie

Joachim Funke ist nicht nur durch mehr oder weniger direkte Tradition Denkpsychologe, er bekennt sich auch wiederholt und in verschiedenen Phasen seines Forscherlebens dazu: „Von Anbeginn der Arbeiten zum komplexen Problemlösen ging es ...um eine inhaltliche Erweiterung des Gegenstandsbereichs der Denkpsychologie“ (Funke, 2003b, S. 137), um „die Fortentwicklung der denkpsychologischen Forschungsbemühungen“ (Funke, 1986, S. 84). Genauer gesagt, Funke versteht die Psychologie des Problemlösens als eine Erneuerung der Denkpsychologie, die zu den älteren Formen der Psychologie des Denkens (Assoziationismus, Gestalttheorie, Informationsverarbeitung, Handlungstheorie, Evolutionspsychologie; siehe Funke, 2022) hinzutritt.

Als Pionier dieser neuen Denkpsychologie bezeichnet Funke wiederholt Dietrich Dörner, dessen Experimente mit der am Computer simulierten Kleinstadt *Lohhausen* „nicht nur für ein neues Untersuchungsfeld der Denkpsychologie [stehen, sondern] zugleich ein Plädoyer gegen die im Wissenschaftsbetrieb vorherrschende ‚analytische Prozedur‘ [sind]“ (Funke, 1986, S. 9). Als Forschungsstradition zeichnet sich die neue Denkpsychologie folglich sowohl dadurch aus, die

höheren kognitiven Funktionen mit neuen Methoden zu untersuchen, als auch dadurch, eigene wissenschaftstheoretische und methodologische Auffassungen zu verteidigen: „Die wohl heftigste Kritik an konventioneller Versuchsplanung und -auswertung im Bereich der Denkpsychologie stammt von Dörner, der in seiner Karikatur eines varianzanalytischen Vorgehens im Wesentlichen die Schwächen schlechter Experimente vorführt, die ohne theoretische Modellannahmen ‚blind‘ Daten erheben und scheitern“ (Funke & Spring, 2006, S. 652). Allerdings bedient sich die Denkpsychologie nicht nur der qualitativen Einzelfallanalysen, d. h. „zunächst Phänomene zu beschreiben und ‚Schmetterlinge zu sammeln‘“ (Funke, 2003b, S. 137), sondern auch „formaler Beschreibungsmittel“ (Funke, 1986, S. 84), um die Struktur des Problemlösens besser zu erfassen.

Neu ist die Denkpsychologie, von der Joachim Funke spricht, weil sie die etablierte Forschungspraxis infrage stellt, insofern „die frühe experimentelle Orientierung der Denkpsychologie auch eine (wohl unbeabsichtigte) Beschränkung des Forschungsgegenstandes mit sich gebracht [hat]: die Reduktion auf Problemstellungen, die im Labor einfach herstellbar und gut messbar sind“ (Funke, 2006, S. 375). Der Fortschritt besteht in der Zuwendung zur Komplexität des Lebens, der Berücksichtigung des Aufbaus von Lebenswelt und Umwelt anstelle ihrer Reduktion auf standardisierte Experimentalsituationen und abstrakte Denksportaufgaben: „Rund ein Jahrhundert lang bestimmten sie die Denkforschung, weil man sie im Labor leicht darbieten und kontrollieren konnte. Heute beschäftigen sich die Denkpsychologen mit komplexen Problemen“ (Funke, 2018, S. 76), also „Problemen in der alltäglichen Umwelt des Menschen“ (Funke, 1986, S. 2). Dass Denkpsychologie hier mehr als Psychologie des Denkens bedeutet, wird daran klar, dass grundsätzliche Überzeugungen, die die Forschung vom Rest der Disziplin unterscheiden, entwickelt und durchaus selbstbewusst vertreten werden:

Hier haben sich Denkpsychologen einem „main stream“ widersetzt, der unter dem Primat der Laborforschung und der dabei nötigen experimentellen Kontrolle die Validität der Fragestellungen aus den Augen verloren hatte. Was die Problemlöseforschung seit rund 25 Jahren mit der Hinwendung zum komplexen Problemlösen bereits in wichtigen

Teilen vollzogen hat, steht anderen Bereichen der Psychologie noch bevor. (Funke, 2003b, S. 254)

Im Gegensatz zur ursprünglichen Denkpsychologie, deren Methoden „nach heutigen Maßstäben als wenig komplex bezeichnet werden [müssen]“ (Funke, 1986, S. 2), orientiert sich die neue Denkpsychologie an der Fülle und Vielfalt menschlicher Verhaltens- und Erlebnisweisen. Dabei kann ihr, so erwägt es Funke, auch die Phänomenologie helfen: „Auch wenn sich dort keine expliziten Bezüge zur Denkpsychologie finden, soll wenigstens festgehalten werden, dass zentrale Konzepte der Phänomenologie wie etwa ‚Bedeutung‘, ‚Intentionalität‘ und ‚Situiertheit‘ sehr wohl Folgen für das problemlösende Denken haben“ (Funke, 2003b, S. 57; hierzu Wendt, 2019). So wird eine neue Denkpsychologie möglich, die, wie Joachim Funke in einem Vortragstitel aus dem Jahr 2006 zu erkennen gibt, als „Experimentelle Denkpsychologie“ (Funke, 2006) zu verstehen ist.

Gleichzeitig war es Joachim Funke besonders in seiner späteren Schaffensphase wichtig, nicht in wissenschaftstheoretischer Reflexion zu verharren, sondern die denkpsychologisch orientierte Problemlöseforschung unter dem Motto „mind in action“ auch in der Praxis zur Geltung zu bringen. Das Wirkungsfeld mit der größten Reichweite war in dieser Hinsicht vermutlich die Beteiligung an den PISA-Studien (*Programme for International Student Assessment*) zwischen 2000 und 2015 (Csapó & Funke, 2017; OECD, 2014), in deren Rahmen zehntausende Schülerinnen und Schüler weltweit die in Heidelberg entwickelten Problemlöseaufgaben bearbeiteten.² Waren die Aufgaben zunächst noch vergleichsweise abstrakt (z. B. das Raumfahrt-Szenario), wurden sie zunehmend realitätsnäher (z. B. Handballtraining oder Pflanzenwachstum), um schließlich auch in den sozialen Bereich in Form des kollaborativen Problemlösens vorzustoßen. Joachim Funks leitender Gedanke dabei war stets, dass systemorientiertes Denken und Problemlösen zu den Schlüsselfertigkeiten des 21. Jahrhunderts gehören (Funke, Fischer, & Holt, 2018).

² Joachim Funke im Interview zu PISA (Funke, 2008).

Zur Festschrift: Ein Anachronismus?

Dem Denkpsychologen Joachim Funke ist diese Festschrift gewidmet. Die Autorinnen und Autoren sind wissenschaftliche Wegbegleiterinnen und Wegbegleiter seiner Biografie und die Vielfalt der Beiträge zeigt, dass Funke mehr als ein Denkpsychologe ist. Sein Werk umfasst von Anbeginn an weitere Themen, seien es Theoretische Psychologie, Umweltpsychologie oder Alkoholismus. Die Geisteshaltung, mit der er sich diesen Gebieten widmet, lässt nichtsdestoweniger das Profil eines Forschers der komplexen Probleme und des Denkens erkennen. Wenn das Motto seines Blogs „not always problem solving“ ist, bestätigt die Ausnahme die Regel.

Dass Funkes Lebenswerk im Horizont einer wissenschaftlichen Tradition steht, bereichert es nur an Bedeutung. Die Bezüge verdeutlichen, welche Forschungsfragen im Hintergrund der Einzelbeiträge stehen, und integrieren das Gesamtwerk. Eine Festschrift ist ein würdiger Ort, um diese Zusammenhänge zu reflektieren. Aus verschiedenen Perspektiven leisten sämtliche Artikel des Bandes ihren Beitrag dazu. Sie lassen sich in drei Gruppen einteilen.

Im ersten Teil sind Texte versammelt, die einen direkten Bezug zu Joachim Funkes Biografie besitzen. Dort finden sich die Rückschau seiner Lehrer, Reminiszenzen seiner Kolleginnen und Kollegen sowie Würdigungen seiner Schülerinnen und Schüler. Der zweite Teil gibt wissenschaftlichen Essays Raum, die die Lage der Problemlöseforschung, des Faches Psychologie, der Universität als Institution oder der Wissenschaft im Allgemeinen kommentieren. Der dritte Teil umfasst die Artikel, die im engeren Sinne den einzelwissenschaftlichen Diskurs führen. Sie entstammen den Forschungsbereichen, mit denen sich Joachim Funke zeitlebens beschäftigt hat.

Die Einladung, zu dieser Festschrift beizutragen, ist auf breite und oftmals begeisterte Zustimmung gestoßen. Vielen war es offensichtlich ein Anliegen, ihre tiefe und lange zurückreichende Verbundenheit mit Joachim Funke zum Ausdruck zu bringen, und auch wir freuen uns, durch diese Festschrift unsere Verbundenheit und Dankbarkeit für seine Begleitung unseres wissenschaftlichen Werdegangs

zeigen zu können. Gelegentlich gab es auch Töne der Überraschung, dass das Format der Festschrift in der Psychologie überhaupt noch existiere – dieses Buch ist wohl ein überzeugender Nachweis dafür, und wir wünschen allen eine inspirierende Lektüre.

Wir bedanken uns bei allen, die zur wissenschaftlichen Qualitätssicherung durch ihre Kommentare beigetragen haben. Dabei handelt es sich neben zahlreichen der Autorinnen und Autoren des Bandes um (in alphabetischer Reihenfolge) Fabian Hutmacher, André Kretzschmer, Sebastian Pannasch, Josh Joseph Ramminger, Gerhard Reese, Bernd Reuschenbach und Jan Rummel. Ferner haben wir dem Verlag der Ruprecht-Karls-Universität, Heidelberg University Publishing (heiUP), zu danken, denn die Zuverlässigkeit, Professionalität, Flexibilität und Freundlichkeit der Mitarbeiterinnen und Mitarbeiter hat dem Projekt eine zuverlässige Grundlage gegeben. Hervorheben möchten wir dabei die wertvolle Unterstützung von Anja Konopka, die unser Projekt stets hilfsbereit zusammen mit Jelena Radosavljević, Johanna Hildenbeutel und mehreren Lektorinnen und Lektoren seit den ersten Schritten Anfang 2022 begleitet hat.

Diese Festschrift erscheint zudem im Angedenken an Benő Csapó, einem langjährigen Forschungspartner und persönlichen Freund von Joachim Funke an der Universität Szeged (Ungarn), der kurz vor Fertigstellung des vorliegenden Bandes überraschend verstarb.

Referenzen

- Csapó, B. & Funke, J. (2017). The development and assessment of problem solving in 21st-century schools. In B. Csapó & J. Funke (Hrsg.), *The nature of problem solving: Using research to inspire 21st century learning* (S. 19–31). OECD Publishing.
- Frijda, N. H., & de Groot, A. (Hrsg.) (1981). *Otto Selz: His contribution to psychology*. De Gruyter.
- Funke, J. (1986). *Komplexes Problemlösen. Bestandsaufnahme und Perspektiven*. Springer.
- Funke, J. (2003a). Kognition und Kommunikation. In D. Weimer & M. Galliker (Hrsg.), *Sprachliche Kommunikation. Ansätze und Perspektiven* (S. 122–124). Asanger.

- Funke, J. (2003b). *Problemlösendes Denken*. Kohlhammer.
- Funke, J. (2006). Komplexes Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen* (S. 375–446). Hogrefe.
- Funke, J. (2008, 9. Januar). *Problemlöse-Diagnostik für PISA* [Video]. YouTube. <https://www.youtube.com/watch?v=4vvm9GYfRVI>
- Funke, J. (2018). Probleme intelligent lösen. In B. Godina (Hrsg.), *Systemisch finale Intelligenz* (S. 75–86). Springer.
- Funke, J. (2022). Denken. In M. A. Wirtz (Hrsg.), *Dorsch. Lexikon der Psychologie*. Hogrefe. <https://dorsch.hogrefe.com/stichwort/denken>
- Funke, J., Fischer, A., & Holt, D. V. (2018). Competencies for complexity: Problem solving in the twenty-first century. In E. Care, P. Griffin, & M. Wilson (Hrsg.), *Assessment and teaching of 21st century skills* (S. 41–53). Springer.
- Funke, J., & Hussy, W. (1977). KOGNI: Ein Computerprogramm für die Berechnung verschiedener Maße zur Erfassung kognitiver Komplexität. *Trierer Psychologische Berichte*, 4(10), 1–26.
- Funke, J., & Spering, M. (2006). Methoden der Denk- und Problemlöseforschung. In J. Funke (Hrsg.), *Denken und Problemlösen* (S. 648–744). Hogrefe.
- Hussy, W. (1973). *Informationsverarbeitung als reduzierender und generierender Prozess. Untersuchungen zu einem Modell der Verarbeitung von stochastisch – ergodischen Zeichenfolgen* [Dissertation]. Universität Trier.
- Hussy, W. (1984). *Denkpsychologie. Ein Lehrbuch: Bd 1. Geschichte, Begriffs- und Problemlöseforschung, Intelligenz*. Kohlhammer.
- Newell, A. (1967). Without benefit of computer. *Science*, 156(3776), 803–804.
- Newell, A., & Simon, H. (1972). *Human problem solving*. Prentice-Hall.
- OECD (2014). *PISA 2012 results: Bd. 5. Creative problem solving: Students' skills in tackling real-life problems*. OECD Publishing.
- Pongratz, L. (1967). *Problemgeschichte der Psychologie*. Francke.
- Wendt, A. N. (2019). Lösung oder Einfall? Über die verlorenen Spuren der Phänomenologie in der Denkpsychologie. In T. Kessel (Hrsg.), *Philosophische Psychologie um 1900* (S. 189–214). Springer.

Teil I

Biografisches

Joachim Funke: 15 Jahre Wegbegleiter

JÜRGEN BREDENKAMP

Rheinische Friedrich-Wilhelms-Universität Bonn

Abstract. Das folgende Kapitel bezieht sich hauptsächlich auf fünfzehn gemeinsam verbrachte Jahre an den Universitäten in Trier und Bonn. Joachim Funke wird vor allem als Wissenschaftler gewürdigt, der das theoriearme Gebiet des komplexen Problemlösens durch formale Modelle bereichert hat, die sich experimentell prüfen lassen. Zusätzlich sind aber auch persönliche Kontakte des Autors zu Joachim Funke Gegenstand dieses Kapitels.

Zum 1. August 1980 habe ich den Ruf auf einen Lehrstuhl für Allgemeine Psychologie an der Universität Trier angenommen, der die Möglichkeit zur Besetzung einiger Mitarbeiterstellen bot. Die meisten Stellen waren für Mitarbeiter und nun diplomierte studentische Hilfskräfte aus Göttingen, wo ich vor meiner Rufannahme tätig war, vorgesehen (Edgar Erdfelder, Silvia Mecklenbräuer, Werner Wippich). Es blieb noch eine Stelle, die zu besetzen war. In einem Forschungsprojekt, das von Dr. Walter Hussy geleitet wurde, der ebenfalls zum Lehrstuhl, den ich gerade übernommen hatte, gehörte, war als eben diplomierter wissenschaftlicher Mitarbeiter ein junger Mann namens Joachim Funke beschäftigt, mit dem ich schnell ins Gespräch kam. Seine Interessen und Kompetenzen lagen in einem Bereich der Allgemeinen Psychologie, der durch die Mitarbeiter aus Göttingen kaum abgedeckt war, nämlich in der Denkpsychologie mit dem Schwerpunkt des Problemlösens. Was mir sofort gefiel, war Joachim Funkes Interesse an der

Theorienbildung mittels formaler Modelle in diesem Bereich. Aber leider schien es keine Möglichkeit zu geben, ihn als Mitarbeiter zu gewinnen, hatte er doch Professor Aebli zugesagt, bei ihm in Bern eine Stelle zu übernehmen. Aebli (1980) hatte gerade den ersten Band seines umfassenden Werkes *Denken: Das Ordnen des Tuns* veröffentlicht, dem ein Jahr später der zweite Band folgte (Aebli, 1981), und eine Stelle bei ihm wäre für Joachim Funke sicherlich ideal gewesen. Dennoch habe ich ihm für Trier ein Stellenangebot unterbreitet. Dass er dieses letztlich angenommen hat, ist wohl auch Edgar Erdfelder zu verdanken, der aus Göttingen mit nach Trier gekommen und hochgradig an der Entwicklung formaler Modelle interessiert war. Beide haben sich schnell miteinander angefreundet.

Mit der Übernahme der noch freien Stelle durch Joachim Funke begann eine 15 Jahre dauernde Zeit, die er als Mitarbeiter an meinem Lehrstuhl verbrachte. In dem ersten, 2 ½ Jahre dauernden Trierer Zeitabschnitt hat Joachim Funke seine Dissertation fertig gestellt, die dem komplexen Problemlösen gewidmet war (Funke, 1984). Beeindruckt hat mich vor allem sein Ansatz, das Problemlösen als Konstruktion kausaler Modelle zu konzipieren. Nach einer kritischen Würdigung des Stands der Forschung im Bereich des Lösens computersimulierter Probleme schildert Joachim Funke in seiner Dissertation, wie sich derartige Probleme mittels formaler Kausalmodelle beschreiben lassen. Weiterhin werden Annahmen über die mentale Repräsentation derartiger Probleme getroffen. Anhand eines fiktiven ökologischen Problems werden in der Dissertation empirische Tests entwickelt, an deren Erweiterung Joachim Funke später intensiv gearbeitet hat. Die in seiner Dissertation geschilderte Theorienbildung ging weit über das hinaus, was damals zum Thema *komplexes Problemlösen* publiziert worden war. Das Promotionsverfahren wurde mit der Disputation zu einem Zeitpunkt abgeschlossen, als ich nach meiner Rufannahme bereits in Bonn tätig war. Unter anderem hatte ich Joachim Funke und Edgar Erdfelder Stellen in Bonn angeboten, und beide haben dieses Angebot, trotz ihrer Bedenken gegen einen Wechsel nach Bonn, schon im Wintersemester 1984/85 angenommen. So kam es, dass Joachim Funke bereits vor seiner Trierer Disputation Lehrveranstaltungen in Bonn anbot.

Die gemeinsame Zeit in Bonn dauerte dann noch einmal 12 ½ Jahre. In dieser Zeit hat Joachim Funke in verschiedenen Forschungsprojekten die theoretischen und experimentellen Analysen im Bereich des komplexen Problemlösens weiter vorangetrieben. 1990 folgte die Habilitation mit einer Arbeit zu eben diesem Thema, die modifiziert mit dem Titel *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung* im Springer-Verlag publiziert wurde (Funke, 1992). Nach der Habilitation war Joachim Funke noch einige Jahre als Hochschuldozent tätig und hat einen Lehrstuhl in Greifswald vertreten, bevor er dann 1997 den Ruf nach Heidelberg angenommen hat. Ich erinnere mich, dass er schon 1986 anlässlich einer gemeinsamen Reise zum Kongress der Deutschen Gesellschaft für Psychologie nach Heidelberg von dieser Stadt geschwärmt hat, sodass ihm mit seiner Berufung dorthin sicherlich ein Herzenswunsch erfüllt wurde.

Bei dem Rückblick auf die gemeinsame Zeit fällt mir wieder auf, dass Joachim Funke und ich kaum gemeinsam publiziert haben. Unsere inhaltlichen Schwerpunkte lagen wohl zu weit auseinander. Joachim Funkes Forschungsinteressen lagen neben dem Alkoholismus im Bereich des Lösens komplexer computersimulierter Probleme, während andere Teilgebiete der Allgemeinen Psychologie meine Aufmerksamkeit auf sich zogen. Gemeinsam war uns aber die Bemühung um eine theoriegeleitete experimentelle Überprüfung unserer Hypothesen, und hier hat Joachim Funke in einem Feld mit Theoriendefizit Beachtliches geleistet. Zum Abschluss seines Lehrbuches über das problemlösende Denken konstatiert er:

Es ist eine Sache, interessante Phänomene aufzuzeigen. Dies ist durch Untersuchungen zum komplexen Problemlösen eindrücklich geleistet worden ... Es ist eine andere Sache, die theoretische Aufarbeitung dieser Phänomene zu leisten. Dies bedeutet nicht nur die Entwicklung eines Netzwerks präziser Begriffe, sondern vor allem deren Verknüpfung zu empirisch gehaltvollen und damit testbaren Hypothesen und die Integration einzelner derartiger Hypothesen in das, was wir eine Theorie problemlösenden Denkens nennen wollen. (Funke, 2003, S. 258)

Dieses Zitat verdeutlicht die Kontinuität seiner schon in der Dissertation sichtbaren Bemühungen um die Entwicklung prüfbarer Hypothesen und deren Tests im Gebiet des komplexen Problemlösens eindrucklich.

Angesichts unserer unterschiedlichen inhaltlichen Schwerpunkte verwundert es vielleicht nicht, dass eine der nur zwei gemeinsamen Publikationen dem Bericht über die Situation der Psychologie in den neuen Bundesländern gewidmet war (Bredenkamp & Funke, 1996). Der damalige Herausgeber der Zeitschrift *World Psychology* Professor U. P. Gielen, hatte mich als ehemaligen Präsidenten der Deutschen Gesellschaft für Psychologie, der mit der Neustrukturierung der Psychologie in Ostdeutschland befasst war (Bredenkamp, 1993), um einen Artikel für seine Zeitschrift zu diesem Thema gebeten, den wir dann gemeinsam verfasst haben. Meine Erinnerungen an die gemeinsame Zeit in Bonn beziehen sich aber auch auf andere Begebenheiten wie die vielen Diskussionen über die Computer-Ausstattung unserer Abteilung. Joachim Funke favorisierte den Macintosh von Apple, und, soweit ich mich richtig erinnere, sind wir seinen Vorschlägen bei der Anschaffung der Rechner gefolgt.

Auch außerhalb des Instituts haben wir viele persönliche Kontakte gehabt, die sogar seinen Sohn Pierre einbezogen, den meine Frau und ich eine Zeitlang an manchen Tagen als Gast bei uns zum Mittagessen hatten. Gut im Gedächtnis geblieben ist mir auch, wie wir unser Haus, das wir im Zuge eines Umzugs vor Beendigung der Mietzeit verlassen hatten, einem polnischen Kollegen von Joachim Funke, der, wenn ich mich recht erinnere, Informatiker war, für ein paar Wochen als Quartier anbieten konnten. Insgesamt waren die gemeinsamen Bonner Jahre eine Zeit intensiver Kontakte mit Joachim Funke und den anderen Mitarbeiterinnen und Mitarbeitern innerhalb und auch außerhalb des Psychologischen Instituts trotz meiner fehlenden Begeisterung für den Rheinischen Karneval, der von ihnen ausgelassen gefeiert wurde. Meine Einstellung zum Karneval ist wohl angesichts meiner Herkunft aus dem protestantischen Norden ohne Versuche der Missionierung toleriert worden.

So häufig wie früher sehen wir uns heute natürlich nicht mehr. Angesichts des zunehmenden Alters müssen immer weiter werdende Strecken zurückgelegt

werden, um persönliche Treffen zu realisieren. Aber die Kontakte persistieren in anderer Form. So erhalte ich regelmäßig noch Postkarten von gemeinsamen Wanderungen und Fahrradtouren, die Joachim Funke und Frau mit Edgar Erdfelder und Frau unternehmen, und ab und zu gibt es auch noch einmal eine Feier, auf der wir uns wiedersehen. Auf einer dieser Feiern habe ich mich schließlich mit Joachim Funke und anderen ehemaligen Mitarbeitern geduzt und damit einen Schritt gemacht, den ich auch schon früher hätte tun können, wenn dafür nur Sympathie und freundschaftliche Verbundenheit eine Rolle gespielt hätten. Diese haben trotz mancher Persönlichkeitsunterschiede natürlich auch mein Verhältnis zu Joachim Funke bestimmt. Aber mit der Vollendung des 70. Lebensjahres war für mich ein Zeitpunkt erreicht, der die Befreiung von der Verpflichtung versprach, die Leistungen dieser Kolleginnen und Kollegen in Gutachten beurteilen zu müssen, und so konnten wir zum *Du* wechseln.

Die zahlreichen Publikationen Joachim Funkes können hier nicht alle erwähnt werden. Das gilt auch für seine unglaublich vielfältigen Aktivitäten. Hingewiesen sei aber auf seinen Blog, der über das Heidelberger Institut für Psychologie informiert, um das er sich mit seinen Interviews von Personen, die einen Bezug zu diesem Institut haben, verdient gemacht hat. Diese Interviews (*oral history*) leisten einen Beitrag zur jüngeren Geschichte des Heidelberger Instituts für Psychologie. Erwähnenswert ist an dieser Stelle auch, dass infolge seiner gemeinsam mit Frau Prof. Guski-Leinwand ergriffenen Initiative vor Kurzem eine Wundt-Büste auf dem Hof des Psychologischen Instituts aufgestellt werden konnte. Damit wurde die Ehrung einer für unsere Wissenschaft berühmten Persönlichkeit vollzogen, die vor ihrer Leipziger Zeit viele Jahre als Assistent von Hermann von Helmholtz, Privatdozent und außerordentlicher Professor für Medizinische Psychologie und Anthropologie in Heidelberg gelehrt und geforscht hat (Lamberti, 1995).

In dem Abschnitt „Würde durch Arbeit“ seines Buches über die Vielfalt der menschlichen Würde behandelt Bieri (2013) auch die Anerkennung als Möglichkeit, sich in der Wertschätzung der anderen zu spiegeln. Mit einer Festschrift zu seinen Ehren wird Joachim diese Wertschätzung der Herausgeber und aller Autoren verdienstermaßen zuteil. Joachims Aktivitäten, Auszeichnungen und Ehrun-

gen vermitteln mir die Sicherheit, mich vor mehr als 40 Jahren bei der Besetzung der Trierer Mitarbeiterstelle nicht getäuscht zu haben. Er hat sogar mehr als das gehalten, was ich mir damals von ihm versprach. Zu Beginn seines achten Lebensjahrzehnts möge gelten, was der Meister des Glasperlenspiels Josef Knecht so ausdrückte: *Und jedem Anfang wohnt ein Zauber inne* (Hesse, 2003).

Referenzen

- Aebli, H. (1980). *Denken: Das Ordnen des Tuns. Bd. 1: Kognitive Aspekte der Handlungstheorie*. Klett.
- Aebli, H. (1981). *Denken: Das Ordnen des Tuns. Bd. 2: Denkprozesse*. Klett.
- Bieri, P. (2013). *Eine Art zu leben. Über die Vielfalt menschlicher Würde*. Carl Hanser.
- Bredenkamp, J. (1993). Zur Lage der Psychologie in den neuen Bundesländern. In L. Montada (Hrsg.), *Bericht über den 38. Kongreß der Deutschen Gesellschaft für Psychologie in Trier 1992* (S. 17–29). Hogrefe.
- Bredenkamp, J., & Funke, J. (1996). A historical change: Psychology in Eastern Germany after unification. *World Psychology*, 2, 103–115. <https://doi.org/10.11588/heidok.00008210>
- Funke, J. (1984). *Komplexes Problemlösen: Kritische Bestandsaufnahme und weiterführende Perspektiven* [Dissertation]. Universität Trier.
- Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung*. Springer.
- Funke, J. (2003). *Problemlösendes Denken*. Kohlhammer.
- Hesse, H. (2003). *Das Glasperlenspiel. Jubiläumsausgabe zum hundertsten Geburtstag von Hermann Hesse*. Suhrkamp.
- Lamberti, G. (1995). *Wilhelm Maximilian Wundt (1832–1920): Leben, Werk und Persönlichkeit in Bildern und Texten*. Deutscher Psychologen Verlag.

Joachim Funke verarbeitet Informationen

ANNEMARIE FRITZ, WALTER HUSSY UND DAVID TOBINSKI

Universität Duisburg-Essen

Als uns die Anfrage für einen Beitrag zu Deiner Festschrift erreichte, hatten Walter, David und ich unabhängig voneinander sofort die Idee, einen solchen Beitrag gemeinsam zu schreiben. Schließlich haben wir alle drei, wenngleich zuweilen zu unterschiedlichen Zeiten und mit verschiedener Intensität, Deinen Lebensweg ein Stück weit privat wie wissenschaftlich begleiten dürfen. Das Eintauchen in unsere gemeinsame Vergangenheit, das Erinnern an die Themen und Thesen, Dialoge und Dispute hat uns beim Schreiben viel Freude bereitet. Wir hoffen, dass Dir diese kleine Hommage gefällt, eine Hommage an einen aus unserer Sicht vielseitigen und kreativen Wissenschaftler, aber auch an einen herzlichen und humorvollen Menschen. Für uns wurde dieser Text auch zu einer Würdigung prägender Zeiten und spannender Erfahrungen, die uns alle vier wohl nachhaltig (wissenschaftlich und menschlich) (mit-)geformt haben. Daher ist dieser Text sehr bewusst als privat-persönliche Reminiszenz gestaltet und nimmt am Rand wissenschaftliche Thematiken auf.

Die Uni Trier als Wegbereiter eines lebenslangen Forschungsthemas

Annemarie Fritz

Wir hatten uns Trier nicht ausgesucht, eher hatte Trier uns gefunden. In der Person von Klaus Schneewind, der als Lehrbeauftragter einige Seminare an der Uni Basel anbot, hatten wir Einblick erhalten in eine moderne, naturwissenschaftlich orientierte Psychologie und mit seiner Unterstützung wechselten wir zum Sommersemester 1975 von der altherwürdigen (aber durchaus auch etwas verstaubten) Alma Mater zur jungen, gerade gegründeten Universität in Trier. Dieser Wechsel war schicksalhaft, grundlegend für Dein lebenslanges Forschungsinteresse an Denk- und Problemlöseprozessen – denn hier lernten wir in der Vorlesung zum Denken von Walter Hussy die aktuelle kognitive Psychologie kennen, und lasen – quasi druckfrisch – das 1976 erschienene Buch von Dietrich Dörner *Problemlösen als Informationsverarbeitung*.

Die neue Sichtweise ließ Dich kritisch auf die aktuelle Intelligenzstrukturforschung blicken, die sich auf die Bearbeitung einfacher Aufgaben beschränkt und resultatorientiert Denkprozesse außen vor lässt. Deine Überlegungen und Fragen dazu haben Reinhold Scheller im Seminar zur Intelligenz in manche Bedrängnis gebracht.

Aber Du hast auch die Problemlöseforschung selbst nie unkritisch betrachtet. Legendar sind Deine Auseinandersetzungen mit der Dörner-Gruppe, die Du mit scharfer Klinge in Wort und Feder geführt hast. In Erinnerung geblieben ist mir Dein Auftritt auf dem 34. Kongress der Deutschen Gesellschaft für Psychologie in Wien 1984. Du hattest Dein Manuskript in Trier liegen lassen und bist gegen die damals gefeierten Größen der kognitiven Psychologie angetreten.

Deine Kritik bezog sich a) auf die Theoriearmut, mit der die Forschung betrieben wurde, b) auf die fehlende Taxonomie von Problemtypen, c) auf methodische Probleme und – immer wieder beklagt – d) auf die nicht signifikanten Befunde zur Testintelligenz. Mit allen Punkten hast Du Dich auseinandergesetzt und dabei die Geschichte der Problemlöseforschung in Deutschland mitgeschrieben. So hast Du z. B. die Schneiderwerkstatt selbst programmiert, um der Frage nachzugehen,

ob Testintelligenz tatsächlich kein Prädiktor für Problemlösegüte ist. Mit der Einführung des Konstrukts der „operativen Intelligenz“ durch Dörner (1986) wurden hier beide Sichtweisen zusammengeführt. Aber auch auf persönlicher Ebene vertieften die intensiv geführten wissenschaftlichen Auseinandersetzungen den gegenseitigen Respekt und trugen Dietrich Dörner die Ehrendoktorwürde in Heidelberg und Dir den Ehrentitel „mein liebster Feind“ ein.

Zurück zu unseren wissenschaftlichen Anfängen: die Saat der Hussy-Vorlesung ist aufgegangen und hat Früchte getragen. Gemeinsam mit Thomas Kindermann, Michael Walk, Roland Himmelsbach, Marie-Theres Strobel und Norbert Kaiser hatten wir die große Chance, in einem der progressivsten kognitionspsychologischen Forschungsprojekte mitzuarbeiten und damit unsere Diplomarbeiten zu verknüpfen. Insbesondere wir vier (Thomas, Micha, Du und ich), die wir zwei gegenüberliegende Zimmer und einen Kühlschrank teilten, blicken auf eine einzigartige Zeit zurück, die unser wissenschaftliches Arbeiten in jeder Hinsicht geprägt hat.

Die Zeit im EIS-Projekt

Walter Hussy

Hallo, lieber Joachim, willkommen im ‚Club der 70er‘! Ich erlebe diese Zeitspanne schon seit einigen Jahren und kann Dir versichern, dass sie auch viel Spaß, Interessantes und Überraschendes bereithält. Aus dem Gesamtangebot wünsche ich Dir natürlich nur das Positive.

Das offiziell beendete Berufsleben wird Dich vermutlich weiter begleiten, denn von heute auf morgen kann man einen so erfolgreichen akademischen Werdegang nicht so einfach hinter sich lassen. Der Unterschied besteht darin, dass der zuvor zwangsläufig vorhandene Stress mehr und mehr einer selbstgesteuerten Beschäftigung mit präferierten Themenbereichen weicht.

Dabei hilft vielleicht ein kurzer Blick in eine weit zurückliegende Zeit, die uns einen gemeinsamen Abschnitt unseres beruflichen Lebensweges, der auch für mich wegweisend war, bescherte: die Zeit an der Universität Trier.

Du kamst aus Basel an eine neu gegründete Universität, die sich noch im Auf- und Umbau befand. Die Räumlichkeiten der alten Pädagogischen Hochschule bildeten die Startrampe, bald ergänzt durch Neubauten, die die Ansiedlung weiterer Fächer und Fachbereiche ermöglichte. Die Psychologie breitete sich in den alten Gebäuden aus, vom anfänglichen Großraumbüro, von dem ich nicht weiß, ob Du es noch kennengelernt hast, bis zu der großzügigen Raumausstattung des EIS-Projekts, das unsere wissenschaftliche Zusammenarbeit in den 1980er-Jahren prägte.

EIS-Projekt stand für die Erforschung der Entwicklung informationsverarbeitender Strukturen und Prozesse. Es fußte auf dem *SPIV-Modell* (Strukturen und Prozesse menschlicher Informationsverarbeitung) und später den Weiterentwicklungen im *MEKIV* (Rahmenmodell zur elementaren und komplexen menschlichen Informationsverarbeitung) und hatte die Zielsetzung, diese solchermaßen definierten bzw. modellierten Strukturen und Prozesse im Kindes- und Jugendalter vergleichend zu verfolgen.

Einen Schwerpunkt bildete dabei die Frage nach der *Operationalisierung des Konzepts der Informationsverarbeitungsqualität* im Bereich der Gedächtnisprozesse und des Problemlösens, mit dem ich mich hier exemplarisch beschäftigen möchte. Wohlbekannt war die sich steigernde Leistungsfähigkeit im fraglichen Altersbereich, z. B. gemessen an der Anzahl behaltener Items bzw. der Anzahl gelöster Problemstellungen. Aber welche der beteiligten Strukturen und Prozesse war jeweils dafür ausschlaggebend?

Um Antworten auf die Fragestellungen zu erhalten, wurde den Untersuchungsteilnehmern Material vorgelegt, das die Analyse differenzierter Verarbeitungsschritte zuließ. Im Bereich des Lernens, Behaltens und Erinnerns waren es z. B. Wortlisten, die Gruppierungen im Sinne von Chunking ermöglichten. Beim Problemlösen, worauf wir uns im Weiteren konzentrieren wollen, wurden *sequenzielle Vorlagen* gewählt, so z. B. die stochastisch-ergodischen Zeichenreihen, deren Abfolge nicht alleine durch den Zufall, sondern, in unterschiedlichem Ausmaß, durch Gesetzmäßigkeiten bestimmt waren.

Diese Gesetzmäßigkeiten, sogenannte *Superzeichen*, zu erkennen, musste somit das Ziel der Teilnehmer sein, um den Verlauf der Zeichenabfolge möglichst genau *Schritt für Schritt* vorhersagen zu können. Zur Veranschaulichung ein Beispiel mit vier Zeichen (etwa die Buchstaben A, B, C, D):

1. Die Buchstaben unterscheiden sich in der Häufigkeit ihres Auftretens, so tritt A in der Gesamtfolge in 20 %, B in 40 %, C in 10 % und D in 30 % der Fälle auf.
2. Die Buchstaben treten nicht unabhängig voneinander auf. Auf Zeichen A folgt mit einer Wahrscheinlichkeit von 80 % das Zeichen B, zu 15 % das Zeichen C und zu 5 % das Zeichen D. Auf Zeichen B folgt mit 50-prozentiger Wahrscheinlichkeit wieder das Zeichen B usw. Die Zeichenabfolge ist also nicht zufällig, sondern sie besitzt *Redundanz*, die durch das Erkennen der Gesetzmäßigkeiten, sprich der Bildung von *Superzeichen*, erzeugt wird. Je besser den Teilnehmern die Bildung der Superzeichen gelang oder, anders ausgedrückt, sie die Redundanz ausschöpften, desto besser ihre Vorhersageleistung. Hier eine kleine Illustration:

Auf die erste Vorhersage „D“ einer Versuchsperson folgt die Rückmeldung „A“.

Schritt 1: A

Auf die zweite Vorhersage „C“ folgt „C“ usw.

Schritt 2: A C

...

Schritt 5: A C D B D

...

Schritt 10: A C D B D B B A B D

...

Schritt 17: A C D B D B B A B D D A C D B B D

...

Schritt 22: A C D B D B B A B D D A C D B B D B A D B B

usw.

3. Man kann die Schwierigkeit solcher Zeichenfolgen erhöhen, indem zusätzlich *Superzeichen höherer Ordnung* eingebaut werden. Dabei handelt es sich um Gesetzmäßigkeiten in der Zeichenfolge, welche durch *bedingte Wahrscheinlichkeiten* zustande kommen. Man nimmt mehrere Zeichen in den Blick und erkennt im Laufe der Bearbeitung der Abfolge, dass z. B. auf die Folge A, C mit einer bedingten Wahrscheinlichkeit von 75 % D folgt. Diese Erkenntnis (dieser Lernprozess) erhöht die Aussicht auf mehr korrekte Vorhersagen.
4. Dieser Prozess lässt sich mit der Bildung von Superzeichen noch höherer Ordnung fortsetzen. Anstelle von zwei werden nun drei zurückliegende Zeichen in den Blick genommen und es wird erkundet, ob sich damit größere Vorhersageerfolge erzielen lassen. Man berücksichtigt dadurch bedingte Wahrscheinlichkeiten bzw. Superzeichen zweiter Ordnung (etwa auf A, C, D folgt zu 100 % B) usw.

Mit zunehmender Anzahl an bearbeiteten Zeichen bot sich den Teilnehmern die Möglichkeit, diese Gesetzmäßigkeiten zu erkennen und als Grundlage für die Vorhersage des nächsten Zeichens zu nutzen. Die Zahl korrekter Vorhersagen bildet die Verarbeitungsleistung eines Teilnehmers ab. Sie ergibt sich aus dem Vergleich der subjektiven Vorhersagen mit der objektiven Zeichenfolge. Allerdings ist es ein relativ grobes Maß im Vergleich zur *Transinformation*, einem Maß der Informationsübertragung vom Sender auf den Empfänger aus der Informationstheorie. Mit diesem Maß wird ermittelt, wie viele der in der Zeichenfolge enthaltenen Informationen auf die Teilnehmer übertragen werden bzw. im Vorhersageverhalten zum Ausdruck kommen. Ohne auf die informationstheoretischen Grundlagen im

Einzelnen einzugehen (*Transinformations-Komponenten-Analyse*), spiegelt dieses Maß den *Erwartungswert* der Teilnehmer wider, den sie an der jeweiligen Position der Vorhersagesequenz bezüglich der einzelnen Zeichen gebildet haben.

Bezieht man nun diese Ergebnisse auf die jeweiligen Altersgruppen und die verschiedenen möglichen Informationsreduktionsmechanismen, so erhält man die benötigten Hinweise für die Überprüfungen der postulierten Hypothesen zur *Entwicklung der an der Informationsverarbeitung beteiligten Strukturen und Prozesse im beobachteten Altersbereich*.

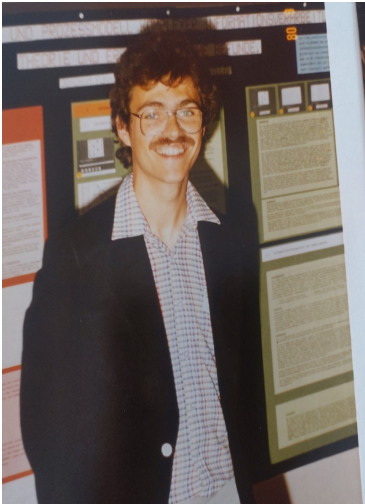
Nach diesem Abstecher in die Arbeitswelt des EIS-Projekts will ich mich noch den persönlichen und sozialen Aspekten im Rahmen der Projektarbeit zuwenden. Bleibende Erinnerungen dürften etwa die parallel zu den wissenschaftlichen Aktivitäten abgelaufenen Freizeitbeschäftigungen hinterlassen haben. Dabei denke ich auch und gerade an die Computer-Fußballturniere, von denen ich im Einzelnen nicht mehr weiß, wie sie organisiert waren und die uns freundschaftlich mit den Mitarbeitern des Kollegen Bredenkamp verbanden, in erster Linie Edgar Erdfelder. Es ist unbestreitbar, dass diese durchaus zeitintensiven Veranstaltungen starke Impulse in die eigentliche Projektarbeit sendeten, nicht nur in wissenschaftlicher, sondern gerade in sozialer Hinsicht, was sich aber – im Gegenzug – sehr positiv auf das gesamte Arbeitsklima und die Projektarbeit auswirkte.

Dieses zeigte sich am Besuch großer Kongresse, aber auch private Feiern und Zusammenkünfte gehörten zu diesem Bild einer geglückten Projektarbeit, die von Dir an zentraler Stelle zu einem erfolgreichen Abschluss gebracht wurde. Einige Bilder aus dieser Zeit (1979–1980) können helfen, die Erinnerungen aufzufrischen (siehe Abb. 1).

Dein Wechsel nach Bonn und meine Berufung nach Köln beendeten unsere räumliche Zusammenarbeit. Für Dich war es ein weiterer Meilenstein auf dem Weg der eingangs erwähnten, so erfolgreichen Karriere, die mit der Berufung auf einen Lehrstuhl für Allgemeine und Theoretische Psychologie an der renommierten Universität Heidelberg ihren krönenden Höhepunkt fand. Die ‚70er‘ sollen Dir die Fortführung Deiner wissenschaftlichen Interessen, aber auch die verstärkte Hinwendung zu privaten Neigungen und Vorlieben ermöglichen. Bleibe gesund!



(a) Posterpräsentation des EIS-Projekts mit Alexander von Eye (DGPs-Kongress 1980 in Zürich)



(b) JoFu, Projektleiter



(c) Die EDV-Ausstattung des EIS-Projekts

Abbildung 1. Zeitdokumente

Zoo-Spiel – ein Transportproblem, bei dem sich wieder Wege kreuzen

Annemarie Fritz

Jahre später haben sich unsere Wege wieder gekreuzt. Ich arbeitete in Köln in der Forschungsgemeinschaft „Das Körperbehinderte Kind“ e.V. und beschäftigte mich mit der Evaluation des Trainingsprogramms „Schule zum Anfassen“. Stark handlungstheoretisch begründet, stand die Vermittlung von Planungskompetenzen im Vordergrund des Trainings. Obwohl *Planen* ein zentrales Konstrukt in der Kognitionswissenschaft ist, fehlten zum damaligen Zeitpunkt valide planungsdiagnostische Testverfahren für jeden Altersbereich. Insbesondere für Kinder lag ein Forschungsdefizit zum Umgang mit komplexen Problemen und zum Erwerb von Planungskompetenzen vor.

In einem ersten Schritt galt es daher zunächst, entsprechende Paradigmen für Kinder zu entwickeln. Angelehnt an das „Missionare und Kannibalen“-Problem (heute wird dies eher als „Flussüberquerungs“-Problem bezeichnet) entstand das „Zoo-Spiel“, bei dem es darum geht, sechs Tiere mit möglichst wenig Fahrten zum Zielort zu transportieren, wobei maximal zwei Tiere gemeinsam transportiert werden dürfen und die Kombination der Tiere bestimmten Regeln unterliegt.

Walter Hussy hatte inzwischen den Ruf an die Universität zu Köln erhalten und war ein Wunschpartner, um gemeinsam Maße für die Güte der Planungsprozesse zu erarbeiten und abzubilden, mit welchen Strategien es Kindern am besten gelingt, die Komplexität des Problemraums einzuschränken. Unsere Arbeiten am Zoo-Spiel wurden von David Tobinski aufgegriffen und weitergeführt (siehe nächster Abschnitt).

Nicht nur Walter und ich fanden wieder zu gemeinsamer Arbeit und stießen dabei auf den Novizen David, auch wir beide entdeckten unser gemeinsames Forschungsthema und bündelten dies in der Herausgabe des Buches *Neue Konzepte und Instrumente zur Planungsdiagnostik* (1995) sowie in einem Symposium auf dem 40. Kongress der Deutschen Gesellschaft für Psychologie in München mit dem Schwerpunktthema „Wissen und Handeln“.

In diese Zeit fällt eine kleine Begebenheit, die heute anekdotischen Charakter hat: 1995 besuchte ich Dich an der Uni in Bonn, um die Herausgabe des Planungsbuchs zu besprechen. Du zeigtest mir die neueste Errungenschaft des psychologischen Instituts – damit könne man Informationen aus aller Welt in Sekundenschnelle abrufen, aber auch miteinander kommunizieren. Man nenne es „Internet“.

Die Zeit danach

David Tobinski

„Frau Fritz Stratmann, kennen Sie das Internet?“ – „Schon lange!“. Wir schreiben das Jahr 2003. Eine kleine Gruppe von Studierenden schart sich um eine Professorin, die kurz zuvor, wie eine Magierin auf einem Kindergeburtstag, ein Krokodil, einen Hund, eine Katze, eine Maus, eine Gans und ein Schaf aus einem Pappkarton gezaubert hat. Einige Studentinnen haben die Holztiere direkt als sehr niedlich bezeichnet und begeistert aufgenommen. Die Professorin greift noch einmal in den Pappkarton und zaubert einen kleinen Wagen hervor. „Darauf fahren die Tiere! Aber immer nur zwei auf einmal!“ Alle knobeln an diesem kleinen Rätsel, einem Interpolationsproblem, wie später noch zu erfahren ist. Einige Studierende verkünden ihre Idee: „In zwei Wochen fährt der Wagen mit den Tieren durch das Internet!“

Ulrich Raatz hat zu dieser Zeit in Duisburg seine Vorlesung zu „Testaufbau und Testanalyse“ gehalten, aber Computer-Based-Assessment war damals noch kein Thema, nicht einmal am Rande. Was hatte diese Gruppe von Studierenden nur dazu getrieben, das Zoo-Spiel (Fritz & Hussy, 2000) ins Internet bringen zu wollen? Das haben sie sich in den darauffolgenden Tagen auch ständig gefragt. Tatsächlich fuhr dank eines selbst geschriebenen Programms zwei Wochen später ein digitaler Wagen mit einer digitalen Gans und einem digitalen Hund über digitale Wege auf einer digitalen grünen Wiese (siehe Abb. 2). „Unglaublich!“, entfuhr es der Professorin. Doch als sie den Wagen zum Krokodil steuern will, fährt dieser einfach mit der Gans und dem Hund weiter im Kreis und weiter und weiter. „Ähm, das ist nur der Teaser! Die Steuerung läuft auch bald! In zwei Wochen!“

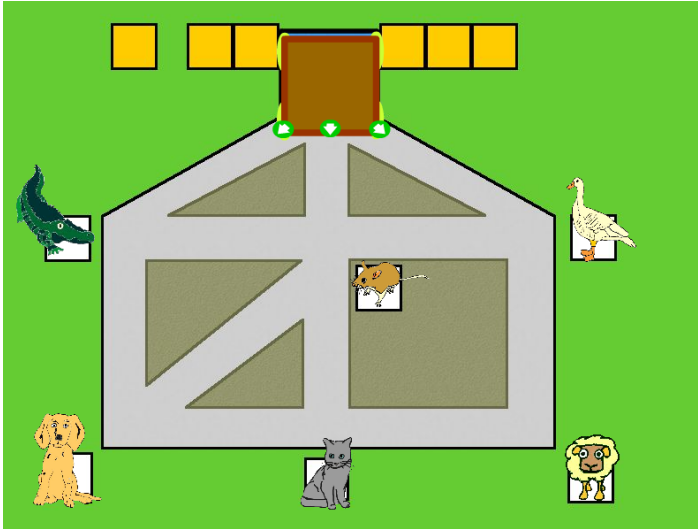


Abbildung 2. *Das Zoo-Spiel*

Das mit den zwei Wochen hat sich im Laufe der folgenden Wochen und Monate verselbstständigt, aber glücklicherweise nahm die Begeisterung nicht ab, sondern zu. Das Lehrbuch *Problemlösendes Denken* von Joachim Funke wurde die Begleitliteratur. „Das ist doch auch der Herausgeber der *Kognitiven Psychologie* von John R. Anderson!“ Er kenne Anderson persönlich, so das damalige Gerücht. Das Lesen von Literatur zum Problemlösen und das Programmieren gerieten in positive Feedback-Schleifen. Der größte Kritikpunkt am Zoo-Spiel wurde dann der Aspekt des One-Trial-Testings. „Zu wenige Items!“, lautete es aus der Ecke der Kolleginnen und Kollegen aus dem Bereich des Large-Scale-Assessments. Aber dennoch sei es „ein interessanter finiter Automat“! Also kam auch noch Automatentheorie zur Testtheorie und zur Theorie des Problemlösens hinzu! Und als dann zudem Karl Popper (1994) postulierte: „Alles Leben ist Problemlösen!“, war zu erkennen: Ein komplexeres Thema als Problemlösen kann man nicht bearbeiten.

Doch dann kam der große Tag. Das Zoo-Spiel sollte vorgestellt werden. Aber nicht in Duisburg, sondern, nein, viel besser, in Heidelberg. Bei Joachim Funke!



Abbildung 3. Der „Funke-Bau“ im Herzen der Universität Duisburg-Essen

Und die Reise nach Heidelberg brachte noch einen Erkenntnisgewinn: Man kann in der Tat ein komplexeres Thema als Problemlösen bearbeiten. Aber wirklich nur eines – komplexes Problemlösen! Es folgten der Weltkongress der Psychologie 2008 in Berlin, 2010 die DGPs in Bremen, 2011 die SWEE und die CEA in Szeged – ein solches Highlight, dass man 2012 dort wieder zusammengekommen ist (mit einer Menge Essener Verstärkung), 2013 die EARLI in München – aus Professor Dr. Joachim Funke wurde mit der Zeit „Jofu“!

Lieber Jofu, während die akademische Psychologie zuweilen den Anschein macht, das Denken in eine Blackbox einsperren zu wollen, so sind die Begegnungen mit Dir immer der lebende Beweis, dass Psychologie *Thinking outside of the box!* sein kann. „Man solle nicht für oder gegen die eigene Zeit, sondern außerhalb ihrer denken.“ (Hösle, 2022). Danke, dass Du die Theoretische Psychologie nicht nur lehrst, sondern auch lebst! Und dass bei den Treffen mit Dir immer außerhalb der eigenen Zeit gedacht wird!

Es wird vielleicht das zweite Gerücht sein, dass Dich die Universität Duisburg-Essen mit dem Funke-Bau, welcher mitten auf dem Campus zwischen Universitätsstraße und Berliner Platz platziert worden ist, aus Heidelberg ins Ruhrgebiet

locken wollte (siehe Abb. 3; aktuell ist dieser scheinbar an eine Mediengruppe untervermietet). Aber 2014 haben wir Dich, unter Mitwirkung von Cleotilde „Coty“ Gonzales von der Carnegie Mellon University, wirklich einmal nach Essen locken können: Dynamic Decision Making!

Die Kritik am One-Trial-Testing der Problemlöseforschung hattest auch Du Dir zu Herzen genommen: Minimal komplexe Systeme für eine Menge Items zu nutzen, mag zwar den Weg in Richtung Item-Response-Theory weisen, aber eignet sich diese Theorie für die empirische Seite der Komplexitätsforschung? Kann Komplexität also überhaupt auf ein „Minimal“ reduziert werden? Ich meine, die Stimme Deines besten und ältesten Feindes (siehe oben) aus der Ferne rufen zu hören. Es klingt zumindest alles auch nach sozialer Komplexität. Zum Glück gibt es „ein[en] Mechanismus der Reduktion sozialer Komplexität“ (Luhmann, 2014): Vertrauen! Vertrauen ist ein Konzept, welches vielleicht seinen Weg in die Modelle der psychologischen Informationsverarbeitungs-forschung finden wird.

Selbst wenn Du der Herausgeber der deutschen Ausgabe der *Kognitiven Psychologie* von John R. Anderson bist, so habe ich mich immer gefragt, zu welchem Paradigma Du Dich letztlich bekennen wirst? Versuchen wir also, wieder außerhalb unserer Zeit zu denken: „Man kann nicht mehr behaupten, dass Dinge ihre Eigenschaften haben innerhalb von Raum und Zeit ohne ihre Beobachtung. ... Nur wie die neue Denkweise aussieht, wissen wir bis heute noch nicht.“ (Uni Wien live, 2023, 24:14–24:26). Auch bei Anton Zeilinger, das Zitat ist der Wiederholung seiner Nobelpreisrede an der Universität Wien entnommen, dreht sich alles um die Information. Wenn der Titel dieses Beitrags „Joachim Funke verarbeitet Informationen“ lautet, so bleibt festzustellen, dass wir uns alle gemeinsam auf den Weg gemacht haben, die Informationsverarbeitung neu zu entdecken. Du hast eine Menge Schülerinnen und Schüler inspiriert und bist für viele ein Mentor unabhängig von universitären Formalitäten geworden. Dafür bin nicht nur ich Dir sehr dankbar! Auch wenn wir die Beobachtung der Beobachtung noch nicht gemeinsam beobachten können (Du bekennst ja in einem Deiner Blog-Beiträge die Liebe zur Farbe Rot), so vertrauen wir bis heute Deinen Gedanken. Und vielleicht ist ein Bekenntnis Deinerseits zu einem Paradigma gar nicht möglich, da

Du der Urheber bist. Also gehört der Titel dieses Kapitels eigentlich erweitert:
„Funkismus: Joachim Funke verarbeitet Informationen!“

Referenzen

- Anderson, J. R. (1983). *The architecture of cognition*. Harvard University Press.
- von Cube, F. (1968). *Kybernetische Grundlagen des Lernens und Lehrens*. Ernst Klett.
- Dörner, D. (1979). *Problemlösen als Informationsverarbeitung* (2. Aufl.). Kohlhammer.
- Dörner, D., Kreuzig, H. W., & Reither, F. (Hrsg.). (1983). *Lohausen: Vom Umgang mit Unbestimmtheit und Komplexität*. Hans Huber.
- Fritz, A., & Hussy, W. (2000). *Das Zoo-Spiel: Ein Test zur Planungsfähigkeit bei Grundschulkindern*. Hogrefe.
- Funke, J. (1990). Systemmerkmale als Determinanten des Umgangs mit dynamischen Systemen. *Sprache & Kognition*, 9(3), 143–154.
- Funke, J. (2003). *Problemlösendes Denken*. Kohlhammer.
- Hösle, V. (2022). *Im Dialog mit Gómez Dávila: Gegenaphorismen, Variationen, Korollarien*. Zu Klampen.
- Hussy, W. (1984). *Denkpsychologie: Bd. 1: Geschichte, Begriffs- und Problemlöseforschung, Intelligenz*. Kohlhammer.
- Hussy, W. (1986). *Denkpsychologie: Bd. 2: Schlußfolgern, Urteilen, Kreativität, Sprache, Entwicklung, Aufmerksamkeit*. Kohlhammer.
- Hussy, W. (1998). *Denken und Problemlösen* (2., überarb. und erw. Aufl.). Kohlhammer
- Mayer, R. E. (1992). *Thinking, problem solving, cognition* (2. Aufl.). Freeman.
- Newell, A., & Simon, H. A. (1961). Computer simulation of human thinking. *Science*, 134(3495), 2011–2017. <https://doi.org/10.1126/science.134.3495.2011>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
- Popper, K. R. (1994). *Alles Leben ist Problemlösen: Über Erkenntnis, Geschichte und Politik*. Piper.
- Uni Wien live (2023, 25. Januar). *Nobelpreis-Vortrag von Anton Zeilinger: Eine Reise durch die wunderbare Welt der Quanten* [Video]. Youtube. <https://www.youtube.com/watch?v=fW4SwcMQYdA&t=1450s>, 24:14 – 24:26.

Zeit mit Jofu: Immer dynamisch, gelegentlich komplex, nie problematisch. Reminiszenzen zweier, die es wissen sollten

JENS F. BECKMANN¹ UND AXEL BUCHNER²

¹Durham University, UK; ²Heinrich-Heine-Universität Düsseldorf

JFB: Sag mal, Axel, kannst Du Dich daran erinnern, wann und unter welchen Umständen Du Jofu das erste Mal getroffen hast?

AB: Oh ja, vergleichsweise genau sogar. Das war im Jahr 1981 in Trier. Ich hatte gerade begonnen zu studieren und mir von einem meiner künftigen Professoren sagen lassen, dass wir doch besser alle wieder nach Hause gehen sollten, weil für all die Leute, die damals Psychologie studierten, nie im Leben genug Arbeitsplätze zur Verfügung stehen würden. Was für eine grandiose Fehleinschätzung! Statt nach Hause bin ich ins Experimentalpraktikum gegangen und dort in einer Gruppe gelandet, die Jofu betreut hat. Das Thema unserer Gruppe war – Überraschung! – das Lösen komplexer Probleme. Ja, so alt ist das Thema in Jofus Forscherbiografie. Technisch waren wir „spitzenmäßig“ – damals eines von Jofus Lieblingswörtern – ausgestattet und hatten sogar einen Tischrechner zur Verfügung. So etwas war zu der Zeit eine ziemlich seltene Erscheinung, die mich schwer beeindruckt hat. Auf dem Rechner hatte Jofu die Schneiderwerkstatt implementiert. Er war davon völlig

begeistert. Seine Begeisterungsfähigkeit im Sinne von ‚sich und andere begeistern‘ ist ja legendär. Mit genau dieser Begeisterung hat er sofort ein paar fundamentale Probleme der Forschung mit der Schneiderwerkstatt und mit anderen dieser sogenannten dynamischen Systeme gesehen, die seinerzeit populär waren. Es war zum Beispiel schick zu behaupten, die Fähigkeit zum Lösen komplexer dynamischer Probleme sei etwas komplett anderes als das, was mit Intelligenztests gemessen wird. Der angebliche wissenschaftliche Beleg für diese steile Behauptung war, dass Maße für die Güte des Lösen komplexer Probleme einerseits und Testintelligenz andererseits vollkommen unkorreliert seien. Jofu hat mit der ihm eigenen Verve schnell die Luft aus diesem Argument gelassen, indem er gezeigt hat, dass damals gängige Ad-hoc-Maße der Problemlösegröße ganz einfach vollkommen unreliabel waren und daher mit einem anderen Maß wie etwa Testintelligenz prinzipiell nicht korrelieren konnten. Es gab noch eine Reihe anderer Probleme der damaligen Forschung zum Umgang mit sogenannten komplexen Problemen, die Jofu nach und nach alle zutage gefördert hat. Das hat nicht nur ihm viel Spaß gemacht und war uns allen eine Lektion. Wir haben gelernt, genau hinzuschauen und uns weder von Modeerscheinungen in der Forschung mitreißen noch von der zur Schau gestellten Autorität etablierter Forschender beeindrucken zu lassen.

Aus dieser Zeit stammt übrigens noch eine Episode, die einiges über Jofus Temperament und seine Grundhaltung verrät. Ich sollte eigentlich nach meinem Abitur Zivildienst leisten, aber es hat sich zunächst keine offizielle Stelle bei mir deswegen gemeldet. Meine Neigung, auf langsame Behörden zu warten, war damals sehr begrenzt. Also habe ich einfach angefangen zu studieren. Prompt trudelte dann während meines zweiten Semesters meine Einberufung ein. Ich habe Jofu davon erzählt. Er fand das empörend und hat sofort eine lange und sehr engagierte Begründung geschrieben, warum so eine Einberufung mitten im Studium großen Schaden anrichten würde - und das hat tatsächlich gewirkt! Ich habe meinen Zivildienst erst nach Abschluss meines Studiums abgeleistet. Auch das ist Jofu meiner Erfahrung nach: Wenn er etwas erfährt, was er grob unvernünftig oder ungerecht findet und er die Möglichkeit sieht, sich zugunsten der Vernunft oder der Gerechtigkeit zu engagieren, dann tut er das auch und praktisch unabhängig

davon, ob er einen persönlichen Nutzen davon hat. Was hatte er schon mit so einem staubigen Studenten wie mir zu tun? Wobei: Man sieht sich im Leben ja immer zweimal und das trifft auf Jofu und mich auch zu. Aber jetzt erstmal zu Dir, Jens: Wie war das bei Dir und Jofu?

JFB: Ich bin ihm mindestens dreimal zum ersten Mal begegnet: Das erste erste Mal war über die Literatur, was sich heute viel einfacher anhört als es damals war. Das lag daran, dass zwischen mir und der jeweiligen Forschungsliteratur ein ‚eiserner Vorhang‘ hing. Dieser eiserne Vorhang sollte nicht nur die Menschen physisch, sondern auch die Ideen voneinander trennen. Beides hat, wie wir alle wissen, nicht ganz funktioniert. So kam ich in den Besitz von Jofus Dissertation und zwar über einen Dozenten, der sich – aus welchen Gründen auch immer – das Privileg ‚verdient‘ hatte, ins sogenannte nicht-sozialistische Ausland zu reisen (NSA, aber das ist eine andere Geschichte). Er brachte unter anderem eine Kopie von Jofus Doktorarbeit von seiner Reise mit, die er auf meinem Schreibtisch in der Uni hat liegen lassen.

Das zweite erste Mal, dass ich Jofu traf, war tatsächlich persönlich. Es muss 1989 oder 1990 gewesen sein, als er das Psychologische Institut der Universität Leipzig besuchte, wo ich damals studierte. Es ist nicht allzu schwer, die Tatsache, dass er dies zu dieser Zeit tat, als Indiz für seine Neugierde, oder nennen wir es ‚Offenheit für Erfahrungen‘, zu sehen. Das ist etwas, was viele seiner Kollegen dieser Generation damals nicht taten, einige taten es erst viel später, und einige haben es nie getan. Aber das wäre wieder eine andere Geschichte, denke ich.

Das dritte erste Mal, dass ich Joachim traf, war, als ich in sein kleines Forschungsteam an der Universität Bonn kam. Und da habe ich auch Dich kennengelernt, erinnerst Du Dich? Das ist mehr als 30 Jahre her, kannst Du das glauben?! Joachim hatte ein DFG-Projekt zur Erforschung des Zusammenhangs zwischen Lernen und Problemlösen bewilligt bekommen. Dieses Projekt war eines der ersten, wenn nicht sogar *das* erste gemeinsame Forschungsprojekt zwischen einem ostdeutschen und einem westdeutschen Forscher nach der Wiedervereinigung, welches die DFG finanzierte. Der glückliche Umstand, dass die Universität Leipzig damals nicht in der Lage war, Gelder ‚aus dem Westen‘ zu verwalten, mit denen mein Gehalt bezahlt

worden wäre, führte zur dynamischen Lösung dieses ‚Unflexibilitätsproblems‘. Mit anderen Worten, ich zog an den Rhein.

AB: Ja, genau, da haben sich unsere Wege gekreuzt. Wer hätte damals gedacht, dass daraus eine dauerhafte Freundschaft zwischen uns werden würde? Aber ich schweife ab – zurück nach Bonn. Jofu hatte auch ein dickes EU-Projekt an Land gezogen und mich angeheuert – da haben er und ich uns dann also zum zweiten Mal getroffen. Man könnte fast sagen, dass sich dadurch Jofus Investition in mein Weiter-Studieren-Können am Ende vielleicht doch noch gelohnt hat. Man müsste ihn mal fragen, ob er das so sieht. An dem EU-Projekt waren neben uns in Bonn vier weitere Arbeitsgruppen beteiligt, unter anderem die von Donald Broadbent in Oxford. Das war spektakulär, ausgesprochen lehrreich und oft unterhaltsam. So manche Anekdote, die Donald Broadbent bei unseren Pub-Besuchen von sich gab, erzähle ich noch heute in meinen Lehrveranstaltungen. Dazu gehört die von einem neuen Flugzeugtyp der britischen Armee, bei dem das fliegende Personal die rechts und links vom Pilotensitz befindlichen Hebel für die Landeklappen und das einfahrbare Fahrwerk – damals beides relativ neue, sehr wichtige Entwicklungen – immer mal wieder verwechselte. Zwei Hebel auseinanderzuhalten ist eigentlich das Gegenteil eines komplexen Problems – einfacher geht es ja gar nicht – und trotzdem kamen Verwechslungen vor. Passierte das bei der Landung, hatte das natürlich fatale Konsequenzen, die wohl kaum jemand in so schöne Worte fassen konnte wie der Brite Donald Broadbent: „This brought the airplane to an embarrassing and expensive collapse.“ Die Tatsache, dass Menschen an scheinbar sehr einfachen Problemen spektakulär scheitern konnten, war für Donald Broadbent ausschlaggebend für sein ausgeprägtes Interesse an angewandter psychologischer Forschung. Er konnte fesselnde Geschichten dazu erzählen, wie sich in seinem Forscherleben angewandte Forschung und Grundlagenforschung gegenseitig befruchtet haben. Das hat mich damals nicht nur schwer beeindruckt, sondern wohl auch geprägt: Angewandte Forschung ist auch meiner Erfahrung nach für die Grundlagenforschung inspirierend. Ich weiß nicht, ob ich ohne diese Prägung durch Donald Broadbent wirklich so viele Projekt zusammen mit der Industrie durchgeführt hätte, gehöre ich doch zu der Generation, die früher alles, nur nicht

„Kapitalistenknecht“ werden wollte. Noch eine Sache muss ich erzählen, auch weil ich weiß, dass Du sie goutieren wirst: Donald Broadbent hat einmal über unser damaliges EU-Projekt bei einem großen öffentlichen Vortrag gesagt, es gehöre zu den schönsten Dingen in seinem Leben, dass er am Ende seiner Karriere mit Leuten aus Ländern zusammenarbeiten darf, die zu Beginn seiner Karriere zu seinen Feinden gehörten. Zum Glück musste er den Brexit nicht mehr erleben...

JFB: Oh, ja! Damit hat sich meine derzeitige Wahlheimat nun wirklich nicht mit Ruhm bekleckert. All dies mutet eher als ein Versuch des „komplexen Problem-Kreierens“ an.

AB: Und wie, für beide Seiten! Wie auch immer, damals war Großbritannien noch Teil der EU, und wir haben uns im Rahmen unseres EU-Projekts um den Erwerb und die Anwendung von Wissen in dynamischen Umgebungen gekümmert. Für uns hieß das: reine Grundlagenforschung. Aus der Zeit stammt die Anwendung des Konzepts finiter Automaten auf die Problemlöseforschung. Nach seinen frühen Erfahrungen mit Ad-hoc-Problemen wie der Schneiderwerkstatt und seiner massiven Kritik daran wollte Jofu vor allem eine solide Basis für diese Forschung schaffen und das hieß: Es mussten erst einmal experimentell kontrollierbare Laborprobleme her, deren Problemräume sich sauber beschreiben und für die sich optimale Lösungen bestimmen ließen. Dazu gehörten Systeme mit Eingabe-Ausgabe-Verknüpfungen auf der Basis linearer Gleichungen und eben solche auf der Basis des Konzepts finiter Automaten.

In späteren Jahren scheinen diese Ideen mit Ansätzen wie MicroDYN und MicroFIN weiter verfolgt und mit dem Gedanken verknüpft worden zu sein, dass man mehrere Messungen von Problemlösegröße haben sollte, die man zu einem halbwegs reliablen Maß aggregieren kann. Das schien mir aus der Ferne betrachtet eine schlaue Idee zu sein, denn es würde ja auch niemand in einem Reaktionszeitexperiment pro Person nur eine Reaktionszeit erheben. Ich habe die weiteren Entwicklungen nicht aktiv verfolgt, mir scheint aber, dass Jofus Zuneigung zu diesem Ansatz später deutlich abgenommen hat. Falls ich das richtig sehe, was waren denn Deiner Einschätzung nach die Gründe dafür?

JFB: Das ist eine interessante Frage, die es sicherlich wert wäre, sie mit Jofu direkt zu diskutieren. An dieser Stelle müssen daher Spekulationen herhalten. Eine Erklärung für Jofus scheinbar schwindendes Interesse am DYNAMIS-Ansatz und an den einfachen finiten Automaten könnte sein, dass ihn die ihm eigene Neugier und Begeisterungsfähigkeit zu anderen, neueren, potenziell spannenderen Themen zog. Du und ich sind ja schließlich auch nicht beim komplexen Problemlösen stehen geblieben.

Ich stimme Dir zu, die MicroDYN-und-MicroFIN-Idee ist oberflächlich betrachtet sehr attraktiv. Sie gestattet zum Beispiel die bequeme Berechnung konventioneller psychometrischer Indizes wie zum Beispiel von Reliabilitätsschätzern und passt in das Schema klassischer psychologischer Leistungsmessung mit multiplen Items innerhalb standardisierter Zeitintervalle. Gleichzeitig verspricht die MicroDYN-und-MicroFIN-Idee, das von jeher mit komplexem Problemlösen assoziierte Versprechen einzulösen, mehr zur Messung intellektueller Leistungsfähigkeit beizutragen als konventionelle Ansätze der Intelligenzmessung. All dies und auch eine gehörige Portion an sicherlich notwendiger Eigenwerbung hat schließlich ermöglicht, dass komplexes Problemlösen Einzug in Large-Scale-Assessments wie zum Beispiel PISA gehalten hat. Das hat zweifelsohne der Komplexen-Problemlöse-Forschung zu einer größeren Aufmerksamkeit verholfen.

Ich habe die Entwicklung zu MicroDYN und MicroFIN gleichzeitig jedoch etwas kritischer wahrgenommen. Ein Grund dafür ist mein Eindruck, dass zum Beispiel die im Kontext des sogenannten DYNAMIS-Ansatzes gewonnenen Erkenntnisse nur unzureichend in der MicroDYN-Umsetzung genutzt wurden. Das erinnert ein wenig an Walter Mischels Einsicht, dass „Psychologists treat other people’s theories like toothbrushes – no self-respecting person wants to use anyone else’s.“

Ein anderes Problem sehe ich im konzeptionell unkritischen Bestreben, konventionellen Reliabilitätskriterien Genüge zu tun. Das Kernelement von Reliabilität ist Stabilität. Komplexes Problemlösen ist jedoch die Fähigkeit, sich mit sich *dynamisch* verändernden Umgebungen erfolgreich auseinanderzusetzen. Das Besondere am komplexen Problemlösen ist das Potenzial, Veränderungen zum

Beispiel in Bezug auf Wissen und Strategieeinsatz *innerhalb* des Problemlösers zu beobachten. Das Interesse an der intraindividuellen Variabilität steht somit im Widerspruch zur traditionellen Vorstellung von Stabilität, bei der die intraindividuelle Variabilität als Fehlervarianz behandelt wird, die es zu minimieren gilt. Kurzum, der MicroDYN-Ansatz birgt die erhebliche Gefahr, dass das Potenzial, welches komplexem Problemlösen eigentlich innewohnt, verloren geht. Bei genauerem Hinschauen muss man sich daher die Frage stellen, ob das Handeln innerhalb von MicroDYN immer noch komplexes Problemlösen ist. Mit anderen Worten, man ist mit dem Risiko konfrontiert, dass psychometrischer Pragmatismus die notwendige Aufmerksamkeit auf eine hinreichende konzeptionelle Fundierung verdrängt. Das Risiko kommt zum Beispiel in der Vagheit dessen zutage, was mit MicroDYN eigentlich gemessen wird: Ist es die Fähigkeit zum logischen Denken, ist es Problemlösefähigkeit, sind es etwa sogenannte ‚21st-Century-Skills‘ oder ist es doch etwas ganz anderes? Ich habe die Vermutung, dass Jofu diesbezüglich selbst Zweifel entwickelt hat. Wäre das tatsächlich der Fall, würde er sich in der klassischen „Die Geister die ich rief“-Position befinden. Ich jedenfalls wäre sehr daran interessiert, mehr über Jofus Ideen zu erfahren, wie Ordnung im „selbstgeschaffenen Chaos“ (siehe Funke, 1991¹) zu schaffen sei.

Die Beantwortung der Frage, ob Handeln innerhalb von MicroDYN tatsächlich noch komplexes Problemlösen ist, hängt meines Erachtens von zwei Faktoren ab. Dies sind erstens ein hinreichendes Verständnis dessen, was wir unter Komplexität verstehen, und zweitens ein mehr oder weniger einheitliches Verständnis dessen, was komplexes Problemlösen selbst ist.

Hinsichtlich des fehlenden Konsensus über die Bedeutung des Komplexitätsbegriffs glaube ich, dass dieses Problem seine Wurzeln in einer vorwiegend datengetriebenen Sichtweise auf komplexes Problemlösen hat. In einer Forschung, die mit Daten beginnt und mit Daten endet, sehen wir die Tendenz, dass der Komplexitätsbegriff durch den Begriff der *Schwierigkeit* ersetzt wird. *Schwierigkeit* ist

¹ Funke, J. (1991). Keine Struktur im (selbstverursachten) Chaos? Erwiderung zum Kommentar von Stefan Strohschneider. *Sprache & Kognition*, 10, 114–118.

jedoch ein psychometrisches Konzept mit begrenzter konzeptioneller und somit erklärender Reichweite. Es informiert uns deskriptiv darüber, dass eine Aufgabe – zum Beispiel ein Intelligenztest-Item oder ein komplexes Problem – von weniger Personen erfolgreich bewältigt wurde als eine andere Aufgabe. Erstere wird dann als schwieriger als letztere deklariert. Ist man an den Gründen für eine solche Beobachtung interessiert, das heißt an einer Erklärung, so endet man in einem tautologischen Argument: Weil mehr Menschen das zweite Problem gelöst haben, ist es das schwierigere der beiden. Die einfache Ersetzung von *Schwierigkeit* durch Komplexität, wie sie in der Forschung zum Lösen komplexer Probleme notorisch oft geschieht, liefert keine Erklärung. Wenn man *Schwierigkeit* und *Komplexität* austauschbar verwendet, missinterpretiert man Beschreibungen als Erklärungen, was letztlich die Theorienentwicklung erschwert. Komplexität spiegelt *ex ante* Überlegungen zu den kognitiven Anforderungen der Aufgabe und den Umständen, unter denen die Aufgabe zu lösen ist, wider. Das macht Komplexität zu einem primär kognitiven Konzept. Schwierigkeit hingegen ist erfahrungsbezogen, personengebunden und per Definition statistisch.

Das zweite Problem, nämlich die verbleibende Vagheit hinsichtlich dessen, was komplexes Problemlösen eigentlich ist, scheint eine Konsequenz des zuvor genannten zu sein. Wie sich beim näheren Hinschauen herausstellt, ist die Antwort auf die Frage, was mit komplexem Problemlösen gemeint ist, ein größeres Problem als von einigen vermutet. Die Bezeichnung *komplexes Problemlösen* wurde in vielen, ganz unterschiedlichen Zusammenhängen und für verschiedene Zwecke verwendet. So wurde *komplexes Problemlösen* beispielsweise zur Bezeichnung eines Forschungsparadigmas verwendet, mit dem psychologische Konzepte wie Informationsverarbeitung, Lernen, Entscheidungsfindung, kausales Denken, Wissenserwerb, Strategieverwendung und anderes untersucht werden. *Komplexes Problemlösen* wurde auch als Bezeichnung für eine vorwiegend kognitive Fähigkeit verwendet, zum Beispiel ‚komplexe Problemlösungsfähigkeit‘. Eine andere, wenn auch wenig informative Verwendung von *komplexem Problemlösen* wäre die Bezeichnung für eine Klasse von beobachteten Verhaltensweisen, die Individuen zeigen, wenn sie mit computersimulierten Mikrowelten konfrontiert

werden. Mit anderen Worten, der Begriff *komplexes Problemlösen* wird verwendet zur Beschreibung einer Forschungsmethodik, zur Beschreibung latenter psychologischer Konstrukte und zur Beschreibung beobachtbaren Verhaltens. Dies sind nur einige wenige Beispiele für die Vielfalt der Verwendung des Begriffs ‚komplexes Problemlösen‘. Diese Vielfalt und die daraus resultierende Bedeutungsunklarheit stellt eine erhebliche Herausforderung für die Beurteilung dessen dar, was mit komplexem Problemlösen gemessen wird.

Schaut man sich die empirische Forschung zum komplexen Problemlösen an, so könnte man den Eindruck gewinnen, dass komplexes Problemlösen je nach den gefundenen Korrelationen zu den Ergebnissen anderer kognitiver Leistungsmaße wie zum Beispiel schlussfolgerndes Denken oder Lernfähigkeit zu einer Fertigkeit ‚eingengt oder herabgestuft‘ oder zu einer Kompetenz ‚erweitert oder aufgewertet‘ wird – oder irgendetwas dazwischen. Eine solche konzeptionelle Flexibilität ist ein Nebeneffekt einer vorwiegend operationalen Definition von komplexem Problemlösen. Das heißt, komplexes Problemlösen ist das, was man mit Aufgabenstellungen misst, die als Komplexe-Problemlöse-Aufgabenstellungen deklariert sind. Die Behandlung von *Fähigkeiten*, *Fertigkeiten* und *Kompetenzen* als Synonyme, die viel zu oft vorkommt, verschärft das Problem der Vagheit noch. Oder, wie man so schön zu sagen pflegt: „What happens in vagueness, stays in vagueness.“

Natürlich gibt es Versuche, das Definitionsproblem konzeptionell besser in den Griff zu bekommen. In einem neueren Versuch, sich in der komplexen Landschaft der Forschung zum komplexen Problemlösen zu orientieren, hat Jofu zusammen mit Dietrich Dörner vorgeschlagen, dass der Begriff des komplexen Problemlösens für den Umgang mit schlecht definierten Problemen reserviert werden sollte. Darin liegt eine gewisse Ironie, denn die Forschung zum komplexen Problemlösen scheint selbst ziemlich schlecht definiert zu sein. Dies kreiert das Problem, dass jeder Versuch, komplexes Problemlösen zu definieren, es zu nicht-komplexem Problemlösen machen würde.

AB: Oha, Jens, das klingt aber sehr nach einer Catch-22-Situation!

JFB: Ja, das ist eine sehr zutreffende Analogie. Die Frage ist letztlich, ob das Kriterium der Undefinierbarkeit ein sinnvolles Element einer Definition sein sollte. Aber das ist ja noch nicht alles. Hinzu kommt, dass die Unterscheidung zwischen ‚gut definiert‘ und ‚schlecht definiert‘ keine Dichotomie ist. Das bringt die Herausforderung mit sich, zu bestimmen, wo auf einer imaginären Definitionsskala ‚gut definiert‘ endet und ‚schlecht definiert‘ beginnt. Hinzu kommt außerdem, dass Probleme, die zuvor als unzureichend definiert galten, mit fortschreitendem konzeptionellem Verständnis immer besser definiert werden können. Ist es nicht seltsam, dass wir nach über 40 Jahren und trotz Jofus zahlreicher und substanzieller Beiträge immer noch nicht genau sagen können, was komplexes Problemlösen ist?

AB: Das ist in der Tat sehr seltsam, aber hey, wenn die menschliche Intelligenz – jedenfalls unsere – hier nicht weiterhilft, dann fragen wir halt die künstliche. Was ich im Sinn habe, ist derzeit noch ziemlich neu und ziemlich aufregend, Du wirst sehen. Es passt auch hierher, denn Jofu war seit jeher von technischen Neuentwicklungen fasziniert. Ich wette, diese hier fasziniert ihn auch. Gehen wir also mal zu ChatGPT (<https://chat.openai.com/>) und bitten um einen kleinen Gefallen.

JFB: Ja, Axel, das klingt spannend. Ich glaube jedoch, dass eine solche Art der Problemlösung neue Probleme aufwirft.

AB: Und wenn schon, schließlich sind wir doch hier, um mehr Probleme zu verursachen als wir lösen.

JFB: I couldn't agree more! Schließlich ist das Hauptziel guter Forschung, nicht in erster Linie Fragen zu beantworten, sondern neue Fragen aufzuwerfen.

AB: Ich sehe, wir sind uns einig. Also los:

> *Write 3000 words about complex problem solving research.*

Complex problem solving is a field of study that focuses on understanding how people solve complex problems and how to improve their problem solving abilities. Complex problems are those that are characterized by a high degree of uncertainty, ambiguity, and dynamic change, and require a range of cognitive, social, and emotional skills to be effectively addressed.

Research in complex problem solving has a long history, with roots in fields such as psychology, education, and computer science. Some of the key areas of research in this field include the development and assessment of problem solving skills, the role of motivation and emotions in problem solving, and the use of technology and other tools to support complex problem solving.

One of the central aims of complex problem solving research is to understand the cognitive processes involved in solving complex problems. This includes understanding how people generate and evaluate potential solutions, how they use prior knowledge and experience to inform their problem solving, and how they adapt to changing circumstances. Researchers have used a variety of methods to study these processes, including experiments, computer simulations, and brain imaging techniques.

One key factor in complex problem solving is the ability to generate creative and novel solutions. Researchers have identified a number of strategies and tactics that people use to generate creative solutions, including brainstorming, lateral thinking, and the use of heuristics and mental models. Some research has also focused on the role of personality and individual differences in creativity, and how these differences may impact problem solving performance.

Another important aspect of complex problem solving is the ability to effectively evaluate and select the best solution from a set of options. This requires the ability to weigh the pros and cons of different solutions, assess their feasibility and impact, and make decisions based on this information. Researchers have studied the decision-making processes involved in complex problem solving, and have identified a number of biases and heuristics that people use to make decisions under uncertainty.

Motivation and emotions also play an important role in complex problem solving. People who are highly motivated and engaged in a problem are more likely to persist in finding a solution, even in the face of challenges and setbacks. Similarly, positive emotions such as curiosity and interest can enhance problem solving performance, while negative emotions such as anxiety and frustration can hinder it. Researchers have studied the ways in which motivation and emotions influence complex problem solving, and

have developed interventions and strategies to improve problem solving performance in these areas.

Technology and other tools can also play a role in complex problem solving. For example, computer-based simulations and modeling tools can help people explore and test different solutions to complex problems. Collaborative technologies such as online forums and social media can facilitate the sharing of ideas and knowledge among problem solvers, and can help people to tap into the collective intelligence of a group. Researchers have studied the ways in which technology and other tools can support complex problem solving, and have developed strategies and guidelines for their effective use.

Overall, complex problem solving is a multi-faceted field of study that involves the integration of many different disciplines and approaches. Through ongoing research, we are learning more about the cognitive, social, and emotional processes involved in solving complex problems, and developing new strategies and tools to support effective problem solving in a range of contexts.

AB: Nicht schlecht für den Anfang, oder? Es sieht fast so aus, als würde die KI in ihrem dritten – oder ist es schon der vierte? – Anlauf endlich wenigstens einen Teil ihrer schon vor Jahrzehnten gegebenen vollmundigen Versprechen einlösen können, auch wenn das Ding offensichtlich noch nicht richtig zählen kann.

JFB: Jetzt warte ich nur noch auf den Tag, an dem mir meine Studierenden Hausarbeiten abliefern, die ChatGPT für sie geschrieben hat. Ich frage mich, ob deren Identifikation so etwas wie „künstliches Problemlösen“ – im Sinne des LöSENS künstlich geschaffener komplexer Probleme – erfordert.

AB: Ein neues Thema für Jofu?

Ein „Funkenschlag“ für die Geschichte der Psychologie

SUSANNE GUSKI-LEINWAND

Fachhochschule Dortmund

Im Wintersemester 1999/2000 hing am Schwarzen Brett im Hintergebäude des Psychologischen Instituts der Ruprecht-Karls-Universität eine Mitteilung, dass ein Thema für eine akademische Qualifikationsarbeit zu *100 Jahre Deutsche Gesellschaft für Psychologie* zu vergeben ist. Absender war Joachim Funke, damals erst relativ kurze Zeit am Lehrstuhl für Allgemeine und Theoretische Psychologie. Mein Diplomstudium der Psychologie hatte ich bereits Anfang September 1999 abgeschlossen und neben meiner Tätigkeit als Lehrbeauftragte für die Absolvent:innen in spe mit „Start in den Beruf“ am PI außerdem den Wunsch, mich noch tiefergehend mit der Disziplingeschichte der Psychologie auseinanderzusetzen. Als damalige Freiberuflerin hielt ich es für wichtig und richtig, bei möglichen Fragen in Beauftragungen auch zum Entstehungszusammenhang psychologischen Wissens Auskunft geben zu können, und auch die Psychologiestudierenden am PI in Heidelberg stellten mir entsprechende Fragen.

Es zeichnet und zeichnete Joachim Funke aus, dass er eigentlich immer ansprechbar war und jedes Klopfen an seiner Tür spontan mit einem herzlichen Willkommen beantwortet wurde, außer es war jemand zur Prüfung oder einem Gespräch im Raum. Für jedes Anliegen hatte er ein offenes Ohr und so auch damals, als ich nach dem Lesen des Aushangs klopfte.

Meine Frage an ihn, ob aus dem Thema ein Dissertationsthema entstehen könnte, nahm er konstruktiv auf und lud für ein weiteres Gespräch und Kontakt zu Carl Friedrich Graumann, seinerzeit als Emeritus noch weiterhin am PI tätig mit Büro im Vordergebäude. Schon bald kamen ein Terminvorschlag und eine Einladung von „JF“.

Wenige Wochen später saßen wir dann zu dritt im Zimmer von JF beieinander, in der Mitte des Tisches ein Schälchen mit Gummibärchen, langjährige, treue Begleiter aus seiner Bonner Zeit und der dortigen Studie. In mehr als einer Stunde umrissen wir zu dritt, wie ein Promotionsthema lauten könnte und was dafür an Einsatz (durch Archivbesuche) und Quellenmaterial (Zeitungsartikel-Recherche, Graue Literatur, bisherige Veröffentlichungen von Carl Friedrich Graumann, Theo Herrmann, Horst Gundlach, Werner Traxel, Helmut Lück und anderen) nötig sei. Ich war dazu bereit und ungemein gespannt auf eine Forschungsarbeit, die mir – für die Psychologie im Allgemeinen damals – eher unübliche Arbeitsplätze in Archiven und anderen Lesesälen bescheren sollte.

Es vergingen knapp drei Jahre mit regelmäßigem Austausch per Mail und persönlich vor Ort, um den Stand der Promotion zu berichten und zu reflektieren. Viele Universitäts-, Staats-, Hauptstaats- und Stadtarchive hatte ich – besonders auch in Ostdeutschland – zu intensiven Recherchen besucht. Und es waren etliche neue Aspekte, die mit den Recherche-Ergebnissen zutage traten. Schwierig wurde es, als ich während eines – eigentlich abschließenden - Lesetermins im Bundesarchiv in Berlin-Lichterfelde auf eine Akte stieß, die einen Brief von Wilhelm Wundts Lehrstuhlnachfolger Felix Krueger an den damaligen Kultusminister Schemm enthielt. Dieser Brief vom April 1933 offenbarte, dass das damalige Leipziger Institut für Psychologie der nationalsozialistischen Politik nahestand und diese befürwortete. Hieraus entstand für mich zentral die Frage, wer die Verantwortung, wer die Verantwortlichkeit für Entwicklungen einer Disziplin trug und wie sich dies zu verschiedenen Zeiten in der Psychologie abbildete. Auf dem Jubiläumskongress der DGPs 2004 trug ich zu dieser Frage einige Dokumente der Verantwortlichkeit zusammen, die dann in Zusammenarbeit mit JF in einen Posterbeitrag mündeten (Guski-Leinwand & Funke, 2004).

Das Recherche-Ergebnis zum Nationalsozialismus weckte in mir den Wunsch, politische Einflüsse und Verflechtungen mit der Psychologie dezidiert zu untersuchen und brachte somit eine Wende in mein zunächst sehr eng umrissenes Promotionsthema. Den beiden Betreuern meiner Dissertation, Joachim Funke und Carl Friedrich Graumann, habe ich es zu verdanken, dass ich mein Thema erweitern konnte und diese bis dato erst einmal nicht erklärliche und auch nicht bekannte historische Tatsache tiefergreifend untersuchte. Anstatt auf dem ursprünglich ausgeschrieben Thema zu bestehen, ermutigten sie mich zu einer Erweiterung der Konzeption meiner Dissertation als Beitrag zur Wissenschaftsforschung (Guski-Leinwand, 2010). Diese umfangreiche Betrachtung der Disziplingeschichte aus Metaperspektiven erforderte zwar dann das Doppelte der ursprünglich geplanten Promotionszeit, hat aber gezeigt, wie wichtig es ist, das Wissen um die Geschichte der Psychologie zu mehren und vor allem auch für eine kritische Professionsreflexion und Fragen der Berufsethik anzuwenden. Darauf hat JF stets hingewiesen und dies als aktuellen und wichtigen Beitrag für unsere Disziplin und für die Profession wertgeschätzt.

Ausgehend von diesem letzten Gedanken erhielt ich nach Beendigung meiner Promotion zur Dr. phil. 2007 die Möglichkeit, zusätzlich einen Lehrauftrag für „Ethik in der Psychologie“ am PI zu übernehmen. Allerdings waren die letzten Monate meiner Promotion getrübt von der Traurigkeit über die schwere Erkrankung von Carl Friedrich Graumann, der nur knapp zwei Tage vor meiner Disputation am 10. August 2007 und damit viel zu früh verstarb. Ähnlich verhielt es sich kurze Zeit später mit dem Tod von Theo Herrmann, der ebenfalls sehr großen Anteil an meinen Untersuchungen nahm.

Ihre Unterstützung, ebenso wie die Ermutigungen von JF, haben mir zu meinem Forschungsschwerpunkt auf dem Gebiet der Geschichte der Psychologie verholfen, dem ich mich mit großer Begeisterung gewidmet habe (Lück & Guski-Leinwand, 2014; Guski-Leinwand, 2017). Mit Beendigung meiner Promotion habe ich mich dann mit den daraus hervorgebrachten Forschungsdesiderata weiter beschäftigt und nach Abschluss meiner Habilitation eine Venia Legendi an der Friedrich-Schiller-Universität für Psychologie und Geschichte der Psychologie erhalten. Als

Privatdozentin im Nebenfach Psychologie lehre ich nun fast seit 10 Jahren „Geschichte und Ethik der Psychologie“ und treffe auf hochinteressierte Studierende, die aus verschiedenen Nachbardisziplinen ihre Fragen an die Psychologie richten und sich inhaltlich engagiert mit ihnen auseinandersetzen.

Mit der Ausschreibung des Bundesministeriums für Bildung und Forschung 2017, Forschungsverbünde zur DDR-Forschung zu bilden, durfte ich zusammen mit weiteren Kolleg:innen von 2019 bis 2023 ein mehr als vierjähriges Teilprojekt zur Rolle und Verwertung der Psychologie in der DDR im Forschungsverbundprojekt „Seelenarbeit im Sozialismus: Psychologie, Psychiatrie und Psychotherapie“ leiten (Strauß et al., 2022; Guski-Leinwand & Maercker, 2022). Unterstützt durch wissenschaftliche Mitarbeiter:innen, haben wir an der Fachhochschule Dortmund unter anderem eine über 7000 Titel umfassende Bibliografie zu psychologischen Veröffentlichungen aus der Zeit der DDR und der Wiedervereinigung zusammengetragen und umfangreiche graue Literatur als Manuskripte aus dem Ministerium für Staatssicherheit der DDR gesichert, die aufzeigen, wie die Politik die Ausdehnung der akademischen Psychologie begrenzte und die Verwertung einzelner Konzepte der Psychologie für Repressionsziele umwidmete (Guski-Leinwand & Nussmann, 2023; Guski-Leinwand, 2024). Die in diesem Zusammenhang an der Juristischen Hochschule des MfS in Potsdam entstandene *Operative Psychologie* zeugt als ein Derivat psychologischer Fachinhalte von einer nicht hinzunehmenden Fehlverwendung psychologischen Fachwissens „gegen den Menschen“, wie nur wenige Jahre nach der Friedlichen Revolution durch den Film *Der schwarze Kasten* von Tamara Trampe und Jochen Feindt aus dem Jahr 1992 bekannt wurde (Trampe & Feindt, 1992).

Nicht zuletzt durch dieses Forschungsprojekt sensibilisiert, sah ich es als wichtig an, den wissenschaftlichen historischen Maßstab, den Wilhelm Wundt als ein Auffinden und nicht Erfinden von (psychologischer) Erkenntnis (Wundt, 1896, S. 317) herausgestellt hat (Guski-Leinwand, 2021), in der Gegenwart auf wahrnehmbare Weise zu verankern: Hier war es wieder die Offenheit von Joachim Funke, der seit 2021 tatkräftig meine Idee, eine Büste Wilhelm Wundts für den Innenhof des Heidelberger Psychologischen Instituts zu spenden, aufnahm und ihr

umfangreich Zeit widmete. Unter Beteiligung vieler universitätsinterner Stellen der Ruprecht-Karls-Universität und externer Stellen wie dem Denkmalamt der Stadt Heidelberg ebnete er der Bronzeplastik des Bildhauers Martin Hintenlang den Weg zur Enthüllung als Stelenobjekt im Frühjahr 2023 unter Beiwohnung vieler Universitätsangehöriger und der Urenkel von Wilhelm Wundt.

Der eingangs genannte Anschlag am Schwarzen Brett lässt sich für mich persönlich, aber sicher auch darüber hinaus, nach mehr als zwei Jahrzehnten eindeutig als ein „Funkenschlag“ für die Geschichte der Psychologie in Heidelberg und in Deutschland erkennen. Mögen die vielfältigen Ergebnisse aus diesem „Funkenschlag“ weitere Kreise ziehen und möge die Geschichte der Psychologie nicht nur als Pflichtbelegung in den Psychotherapie-Studiengängen, sondern auch in den grundständigen Bachelorstudiengängen eine angemessene Verortung erfahren, die als Lehrangebot auch ein Heraus- oder Alleinstellungsmerkmal darstellen könnte. Joachim Funke wünsche ich zur Vollendung seines nun siebten Lebensjahrzehnts eine schöne kommende Zeit, in der er sich auch den anderen Dingen widmen kann, für die sein Herz brennt. Für seine unermüdliche Unterstützung, Bestärkung und Offenheit gegenüber allen Gedanken und Initiativen danke ich ihm zutiefst.

Danksagung. Die persönliche Offenheit und Begeisterungsfähigkeit, die Joachim Funke auszeichnen, stellen für meine akademische Laufbahn eine ganz grundlegende Voraussetzung dar: Ohne seine Bereitschaft zur Betreuung meiner historischen Dissertation am Heidelberger PI wären viele meiner in den Folgejahren stattgefundenen Untersuchungen wohl formal gar nicht durchgeführt worden, weil sie als Forschungsdesiderate nicht erkannt worden wären. Meine Dankbarkeit seiner Offenheit gegenüber bezieht sich hierbei besonders auch auf die Tatsache, dass die Methoden der historischen Forschung nicht seinem präferierten bzw. professionellen Methodenspektrum entsprachen, dies jedoch während der Beratungen zur Dissertation niemals zum strittigen Punkt wurde. Ganz im Gegenteil motivierte Joachim Funke mich immer wieder mit seiner eigenen Haltung, dass das Wissen um die Geschichte der Psychologie „sehr wichtig“ (so sagte er) für unsere Disziplin sei. Insoweit habe ich ihn nicht nur als sehr „methodentolerant“ erlebt, sondern

es hat auch die Disziplin(-geschichte) der Psychologie an methodischem Vorgehen durch „Funkenschläge“ dieser Art gewonnen. Zu meiner gesamten Zeit als Studentin, Lehrbeauftragte und Promovendin an der Ruprecht-Karls-Universität in Heidelberg habe ich „JF“ stets als Professor, Doktorvater und Mentor erlebt, der für alle Belange und Ideen immer eine offene Tür und ein offenes Ohr hatte: Durch seine Nahbarkeit und positive Ausstrahlung konnte ich ihm unumwunden meine Ideen vortragen, bei denen ich niemals Angst haben musste, dass er sie als unmöglich oder undurchführbar einschätzen würde. Immer fand Joachim Funke einen Weg, um das Vorgetragene zu realisieren. Diese verschiedenen Funkenschläge meiner akademischen Laufbahn haben einen nun seit 25 Jahren beschrittenen Weg ausgeleuchtet, von dem ich mir wünsche, dass das Licht, das auch mit seiner Hilfe ins Dunkle mancher Kapitel der Psychologiegeschichte gebracht wurde, einen sonnigen Lichteinfall auf Joachim Funkes nächstes Lebensjahrzehnt wirken möge.

Referenzen

- Guski-Leinwand, S. (2010). *Wissenschaftsforschung zur Genese der Psychologie in Deutschland vom ausgehenden 19. Jahrhundert bis Mitte des 20. Jahrhunderts* [Dissertation, Universität Heidelberg]. heiDOK Heidelberger Dokumentenserver. <https://doi.org/10.11588/heidok.00007667>
- Guski-Leinwand, S. (2017). *Psychologie und Totalitarismus. Die Abwendung vom Humanitätsgedanken in der Psychologie und die Folgen (ca. 1895–1945)*. Peter Lang. <https://doi.org/10.3726/b10520>
- Guski-Leinwand, S. (2021). 1913 – Das Hochjahr der Heterodoxien in der Psychologie: Ein Abriss. In M. Lessau, P. Redl, & H.-C. Riechers (Hrsg.), *Heterodoxe Wissenschaft in der Moderne* (S. 91–102). Brill/Wilhelm Fink. https://doi.org/10.30965/9783846765883_007
- Guski-Leinwand, S. (2024). *Psychologie unter politischem Diktat in der DDR*. Psychosozial-Verlag. <https://doi.org/10.30820/9783837962079>
- Guski-Leinwand, S., & Funke, J. (2004). Dokumente der Verantwortlichkeit: Was lernen wir aus der Geschichte der Psychologie für die Zukunft? Kongress-Posterbeitrag. In T. Rammsayer, S. Grabianowski, & S. Troche (Hrsg.), *44. Kongress der Deutschen*

Gesellschaft für Psychologie. 100 Jahre Deutsche Gesellschaft für Psychologie (S. 297). Pabst Science Publishers.

- Guski-Leinwand, S., & Maercker, A. (2022). Unter politischem Diktat: Psychologie, Psychologische Kriegsführung und Psychopolitik im SED-Staat. In B. Strauß, R. Erices, S. Guski-Leinwand, H. Grabe, & E. Kumbier (Hrsg.), *Seelenarbeit im Sozialismus. Psychologie, Psychiatrie und Psychotherapie in der DDR* (S. 111–138). Psychosozial. <https://doi.org/10.30820/9783837978315>
- Guski-Leinwand, S., & Nussmann, H. D. (2023). *Bibliografie psychologischer Publikationen DDR/Wiedervereinigung. Dortmund*. Fachhochschule Dortmund. <https://doi.org/10.23668/psycharchives.12834>
- Lück, H. E., & Guski-Leinwand, S. (2014). *Geschichte der Psychologie. Schulen, Strömungen, Entwicklungen* (7., vollst. überarbeitete Aufl.). Kohlhammer.
- Strauß, B., Erices, R., Guski-Leinwand, S., & Kumbier, E. (Hrsg.). (2022). *Seelenarbeit im Sozialismus: Psychologie, Psychiatrie und Psychotherapie in der DDR*. Psychosozial. <https://doi.org/10.30820/9783837978315>
- Trampe, T. (Drehbuch, Regie, Interviews), & Feindt, J. (Drehbuch, Regie, Kamera). (1992). *Der Schwarze Kasten* [Dokumentarfilm]. bpb. Bundeszentrale für politische Bildung. Mediathek. <https://www.bpb.de/mediathek/video/346107/der-schwarze-kasten/>
- Wundt, W. (1896). Ueber naiven und kritischen Realismus. *Philosophische Studien*, 12, 307–408.

Teil II

Essays

Drei Jahrzehnte Gummibärenforschung: Entwicklungstrends und neue Ergebnisse

JOCHEN MUSCH¹ UND EDGAR ERDFELDER²

¹Heinrich-Heine-Universität Düsseldorf; ²Universität Mannheim

Mit einem Paukenschlag trat die Gummibärenforschung ins Licht der Öffentlichkeit. Während des Karnevals im Jahr 1995 nahmen Joachim Funke und Heike Gerdes am Psychologischen Institut der Universität Bonn den weltweit ersten Gummibärenserver in Betrieb. Dies geschah zu einer Zeit, als das Internet noch aus kaum mehr als einer Handvoll langsamer Rechner bestand; außer in Bonn war noch an überhaupt keinem psychologischen Institut der gerade wiedervereinigten Bundesrepublik ein WWW-Server in Betrieb. Vielleicht auch deshalb stieß der weltweit erste Gummibären-Preprintserver auf rasch wachsendes Interesse. Rekordabrufzahlen erzielten die Gummibärenforschungsseiten, als sich auch Vertreter etablierter Presseorgane für das elastische Forschungsthema zu interessieren begannen (Knoke, 2009). Im *Internet Archive*, dem Online-Museum für Weltkulturgüter, wurden viele der von dem neuen Forschungsthema inspirierten Arbeiten für die Nachwelt konserviert. Sie sind dort auch heute noch jederzeit abrufbar (Funke & Gerdes, 1995).

Eine zentrale Forschungsrichtung identifizierte dabei die Grundkonstante des Gummibären-Universums, die Zahl 5 („The magical number five: plus or minus

two“, Miller, 1995). Fünf unterschiedliche Farben finden sich in jeder handelsüblichen Gummibärentüte; fünf Extremitäten weist das Gummibärchen inklusive seines Kopfes auf; fünf Gramm wiegt ein Gummibär unter atmosphärischen Normalbedingungen; fünf hoch fünf Bären passen gerade noch in eine Gummibärentüte. Für große Erregung sorgte auch ein Beitrag von Döring (1997) über die sexuellen Fantasien der bunten Bären.

Viele weitere Forschungsarbeiten beschäftigten sich mit den unterschiedlichsten Aspekten des Lebens inner- und außerhalb von Tüten (Funke & Gerdes, 1995). Musch (1997) hat die bis zum Ende des Jahrtausends zum Thema Gummibären erarbeiteten Befunde in einem Übersichtsbeitrag erstmals zusammengefasst.

Seit diesen ursprünglichen Veröffentlichungen sind allerdings fast drei Jahrzehnte vergangen. In dieser Zeit hat sich das Feld der Gummibärenforschung kontinuierlich weiter ausdifferenziert und auch methodisch weiterentwickelt. In einigen Gebieten haben sich dabei überraschend ganz neue Perspektiven eröffnet, und auf Tagungen ist deshalb derzeit häufig Aufbruchstimmung in der noch immer wachsenden Gummibärenforschungsgemeinde zu spüren. Der 70. Geburtstag des Vaters der Gummibärenforschung ist deshalb ein günstiger Zeitpunkt, um das in den letzten Jahren auf diesem wichtigen Gebiet Geleistete zusammenzufassen, es neu zu ordnen und zu bewerten. Dies zu leisten, ist das Ziel der als Überblicksarbeit angelegten ersten Hälfte des vorliegenden Beitrags. Er gliedert die rezente interdisziplinäre Gummibärenforschung zur Systematisierung und besseren Übersicht nicht chronologisch, sondern nach Fachgebieten. In der zweiten Hälfte der Arbeit schließt sich daran ein empirischer Originalbeitrag an, der hier erstmals der Fachöffentlichkeit vorgestellt wird. Er verspricht einen wichtigen ersten Beitrag zur Nutzenmessung bei Gummibären zu leisten und wartet dabei mit einigen überraschenden Befunden auf; insbesondere konnte erstmals das Phänomen der *Nutzenexplosion* farbiger Gummibären mit zunehmender Kardinalität der Gummibärenmenge gezeigt werden. Darüber hinaus fanden sich erste empirische Hinweise darauf, dass der mit zunehmender Anzahl steigende Nutzen der Tütenbewohner auf ihrer visuell wahrgenommenen Schönheit und nicht auf ihrem Geschmack beruht. Es kommt offenbar zu einer Fehlattribution der wahrgenommenen Schönheit auf

den Geschmack der Bären, also zu einer Geschmacksillusion, die im zweiten und empirischen Teil des vorliegenden Beitrags erstmals im Detail nachgewiesen wird.

Geschichte

Im Berichtszeitraum der vorliegenden Übersicht wurden erstmals einige historische Beiträge zur bislang im Dunkeln liegenden Entstehungsgeschichte der Gummibären vorgelegt (zusammenfassend Mommsen, 1995). Schon der ehemalige Kaiser Wilhelm II. – der sonst an der neuen Republik kein gutes Haar ließ – pries im holländischen Exil, wohin er sie sich liefern ließ, die Gummibären von Hans Riegel aus Bonn (dafür steht die Abkürzung „Haribo“) als „das Beste, was die Weimarer Republik je hervorgebracht hat“. Auch dass Albert Einstein und Konrad Adenauer die Gummibären mochten, ist historisch gesichert. Haribo warb für die Bären ab 1935 mit dem Slogan „Haribo macht Kinder froh“; 1962 folgte hierzu die Ergänzung: „und Erwachsene ebenso“. Beide Slogans gehören zu den erfolgreichsten Kampagnen in der Geschichte der Werbewirkungsforschung und führten zu einer stetig zunehmenden Nachfrage. Schon am Ende des vorigen Jahrtausends wurden täglich hundert Millionen Goldbären produziert (Mommsen, 1995).

Medizin

Im Bereich der Chirurgie entwickelten sich Gummibären in den letzten beiden Jahrzehnten unaufhaltsam zu einem wichtigen Werkzeug der medizinischen Forschung. Viel beachtet wurde dabei vor allem der Beitrag von Throckmorton, Ellis und Hayasaki (2004) zum Gummibärkauen als neuer Standardmethode zur Untersuchung von Frakturen im Unterkieferbereich. Die Aspiration von Gummibären durch die Nase wurde in einem aufsehererregenden Fallbericht aus Griechenland als mögliche Ursache für Erstickungsanfälle bei Kleinkindern identifiziert; als einzige zur Behandlung zur Verfügung stehende Interventionsmaßnahme musste eine sofortige Notfallbronchoskopie der Lunge zum Einsatz gebracht werden (Tavladaki et al., 2012). Vor dem Hintergrund derart besorgniserregender Vorfälle muss das bislang noch völlige Fehlen von Beiträgen zu den medikolegalen Aspek-

ten der Behandlung der Nebenwirkungen des Gummibärverzehr als empfindliche Lücke in der Reihe der Beiträge zur Gummibärenforschung angesehen werden.

Psychologie

Wesentliche Fortschritte wurden im Berichtszeitraum vor allem hinsichtlich der differenziellen Psychologie von Tütenbewohnern erzielt. So konnte anhand von DNA-Analysen erstmals nachgewiesen werden, dass alle Gummibären identische Klone von fünf unterschiedlich gefärbten Stammgummibären sind, die von der Firma Haribo in Bonn in den 20er-Jahren des letzten Jahrhunderts hergestellt wurden. Einen weiteren Durchbruch erzielte die weltweit erste Studie mit zwar genetisch identischen, aber in getrennten Tüten aufwachsenden Gummibären. Dabei ergaben sich eindeutige Hinweise auf eine perfekte Heritabilität der Farbgebung bei gleichzeitig großem Umwelteinfluss auf die Basistemperatur der kleinen Bärchen (Bouchard, 1989).

Heftige Kritik entzündete sich unter diagnostischen Gesichtspunkten am Gummibären-Orakel, das bei seiner ersten Veröffentlichung noch auf großes Interesse gestoßen war. Anfangs enthusiastisch berichtete Vorhersageleistungen, die jedoch auf kleinen Stichproben beruhten, konnten nicht repliziert werden; überzogene Aussagen zur prognostischen Validität der verwendeten Typologien brachen letztlich völlig in sich zusammen (Nosek, 2015).

Verkehrsforschung

Die Entwicklung des Gummibär-Positions-Systems (GPS) führte im Berichtszeitraum zu einer Revolution im Bereich der Verkehrsforschung. Nachdem GPS-unterstützte Gummibären sich als zuverlässige Chauffeure selbstfahrender Autos erwiesen, konnte im Fortbewegungssektor erstmals die ersehnte Erreichung der Autonomiestufe 5 vermeldet werden. In der hierzu durchgeführten Begleitforschung wurden Gummibären am Lenkrad als angenehme Gesprächspartner beschrieben, die menschliche Eingriffe in den Lenkprozess vollkommen entbehrlich machten. Der dadurch erzielte Komfortgewinn für die menschlichen Mitfahrer wurde in

den bislang durchgeführten Studien zur Ergonomie selbstfahrender Fahrzeuge einhellig gelobt.

Noch zu früh ist es derzeit für eine abschließende Bewertung der kürzlich vorgestellten These, dass Elon Musk, der Hersteller des ersten gummibärbetriebenen Elektro-Fahrzeugs, selbst ein Gummibär sei. Aufgekommen war diese Vermutung, als Elon Musk aus bis heute ungeklärten Gründen an der Spitze einer „Falcon Heavy“-Rakete einen weißen Gummibären am Steuer eines Tesla Roadster ins All schoss.

Chat-GBT: Die raketenhafte Entwicklung der Gummi-Bären-Transformer

Besonders im Fokus der jüngeren Forschung steht die explosive Entwicklung im Bereich der Gummi-Bären-Transformer (GBT), die auch als „sprechende Gummibären“ Bekanntheit erlangt haben. Sie erweitern Roboterbären mit einer Chatbotfunktion und einem Text-to-Speech-Audiokonverter. Einige der so weiterentwickelten sprechenden Gummibären wurden binnen kürzester Zeit erfolgreiche Influencer auf Online-Plattformen wie Youtube und TikTok. Dabei stellten unabhängige Vergleichsstudien fest, dass die Videobeiträge sprechender Gummibären hinsichtlich ihrer linguistischen Qualität und rhetorischen Performanz bereits heute 95 % des auf den Plattformen vorhandenen Materials substantiell übertreffen.

Intensiv diskutiert wird deshalb derzeit, ob ein Einsatz von Gummi-Bären-Transformern auch im Bildungsbereich möglich und sinnvoll ist. Wie sich gezeigt hat, lieben Kinder die sprechenden Gummibären. Die Frage, welche Effekte der Einsatz bunter Bären in der Schule auf die kindliche Sprachentwicklung hat, wird angesichts der stürmischen Entwicklung und zunehmenden Verbreitung von Gummibären in Bildungseinrichtungen derzeit intensiv in Feldstudien erforscht.

Auch im Hinblick auf die möglichen Auswirkungen auf die Arbeitswelt erscheint derzeit vieles noch unklar. Die Erwartungen gehen dabei weit auseinander. Nicht wenige Beobachter werten sprachgenerierende Gummi-Bären-Transformer als epochalen Zwischenschritt beim Durchbruch hin zur Entwicklung einer allgemeinen

künstlichen Intelligenz. Andere und sehr erfahrene Koryphäen wie Noam Chomsky kritisieren die neuen Systeme hingegen spöttisch als „statistische Papageien“ und „Münder ohne Gehirn“ (Chomsky, 2023).

Kann und wird der *Ursus elasticus* in Zukunft den *Homo sapiens* zunehmend ersetzen? Wird die Arbeit von Menschen in textproduzierenden Branchen zunehmend durch die Gummi-Bären-Transformer der neuesten Generation ersetzbar? Wie steht es um die Bedeutung von Gummibären für die psychosoziale Versorgung (siehe Endepohls & Funke, 2004) und was sind die berufspolitischen Konsequenzen des zunehmenden Einsatzes von künstlicher Bärenempathie für die psychotherapeutische Versorgung? Beruhen die scheinbaren Erfolge GBT-basierter Therapeuten nicht auf denselben anthropomorphisierenden Mechanismen, die auch schon Joseph Weizenbaum im Jahr 1966 bei der Vorstellung seines Eliza-Chatbots an der Urteilsfähigkeit seiner Mitmenschen verzweifeln ließ? Möglicherweise ist gegenüber dem weitgehend unreflektierten und überwiegend ökonomisch motivierten Einsatz von Gummibären in hoch sensiblen Bereichen mehr kritische Skepsis angebracht, als sie derzeit häufig zu beobachten ist.

Rechtswissenschaft

In der juristischen Fachliteratur wurden zuletzt vor allem urheberrechtliche Fragen im Zusammenhang mit der Textproduktion von Gummi-Bären-Transformern diskutiert. Relevant sind diese Fragen auch für künstlich hergestellte Gummibären-Bilder. Die stürmische Entwicklung im Bereich des Prompt-Crafting zwingt die Kunstwelt, Fragen nach Kreativität und Urheberschaft völlig neu zu diskutieren. Bislang galt Kreativität als ein Alleinstellungsmerkmal des *Homo sapiens*. Wer aber ist eigentlich der Künstler, wenn künftig zwar der Mensch noch den Algorithmus vorgibt, der Bär aber den Text und das Bild erschafft? Werden sich der *Homo sapiens* und der *Ursus elasticus* künftig die Urheberschaft teilen? Und auf wessen Konto fließen dann die Tantiemen? Der Austausch der Argumente zu diesen berufspolitisch eminent wichtigen Fragen hat in der juristischen Praxis derzeit gerade erst begonnen.

Weltraumforschung

Kontrovers diskutiert wurde zuletzt die Möglichkeit einer Marsmission mit Gumminauten, die intensiv geprüft wird, seit eine mit menschlichen Astronauten in der Wüste von New Mexiko durchgeführte Simulation der psychosozialen Belastungen bei langen Weltraumflügen desaströs und sehr nachtragend endete. Aufgrund der durch den geplanten Hohmann-Transfer vorgegebenen energetischen Rahmenbedingungen konnte anhand der Umlaufbahnen der beteiligten Himmelskörper sogar bereits der genaue Tag der Ankunft des ersten Gumminauten auf dem Mars errechnet werden. Am 2. August 2048 wird die Landung erfolgen (Walter, 2016).

Quantengummi

Noch immer in den Kinderschuhen steckt seit langer Zeit die dennoch von vielen aufmerksam verfolgte Forschung zur Entwicklung von Quantengummi. Anders als herkömmliche Materie soll Quantengummi viele Zustände im Möglichkeitsraum gleichzeitig annehmen können. Seine Herstellung würde es Gummibären ermöglichen, mithilfe eines selbstinduzierten Kollapses der Wellenfunktion spontan an beliebigen Orten zu materialisieren.

Die vielfältigen hieraus resultierenden Anwendungsmöglichkeiten haben das Interesse an der Forschung zur Quantengummi-Teleportation nie erlahmen lassen. Im allerdings ungewissen Erfolgsfalle hätte die Forschung in diesem Bereich weitreichende Konsequenzen für die Durchführung interstellarer Wochenendreisen (Hawking, 2020).

Eine empirische Untersuchung zum Nutzen von Gummibären¹

Ein weiterer wichtiger Schwerpunkt der Gummibärenforschung war stets auch die Erforschung höherer geistiger Prozesse, wie sie etwa der Kommunikation

¹ Mit diesem Forschungsbericht löst der Zweitautor des vorliegenden Kapitels ein Versprechen ein, das er Joachim Funke 1995 gab – in der ersten Blütezeit der Bonner Gummibärenforschung also. Die Realisierung dieses Forschungsberichts wurde ursprünglich noch für 1995 in Aussicht gestellt,

(Hoepfner, 1997) und dem komplexen Problemlösen (Funke, 1995) zugrunde liegen. Während menschliche Problemlöser im Umgang mit komplexen Systemen gravierende Defizite offenbarten (Dörner, 1989), zeigten Gummibären – zur Überraschung vieler – ganz hervorragende Problemlösefähigkeiten. Insbesondere Stresssymptome und voreilige Schlüsse traten bei Gummibären fast nie auf (Funke, 1995). Gleichwohl wurde den Tütenbewohnern aber ein gewisser Hang zur Lethargie und bisweilen auch ein Mangel an Entschlussfreude attestiert (ebenfalls Funke, 1995).

Vor diesem Hintergrund muss es als erstaunliche Lücke erscheinen, dass entscheidungspsychologische Erkenntnisse bislang selten auf Gummibären angewandt wurden. Diese Forschungslücke dürfte mitverantwortlich dafür sein, dass die Frage, ob eine schweizerische Firma ihren Schokoladenbären „Teddy“ in goldener Folie verpackt vertreiben darf, letztlich vom Bundesgerichtshof entschieden werden musste (Radomsky, 2015). Offenbar befürchtete der klagende (und letztlich im Rechtsstreit unterlegene) rheinländische Produzent der „Goldbären“ wirtschaftliche Nachteile, obwohl dessen Goldbären bekanntlich Gummibären sind, die weder optisch noch geschmacklich Ähnlichkeiten mit „Teddy“ aufweisen.

Der Rechtsstreit hätte vermutlich verhindert werden können, wenn es ein solides empirisches Fundament zum Nutzen von Gummibären gegeben hätte, das auch die Untersuchung von Wechselwirkungen mit vermeintlichen Konkurrenzprodukten erlaubt. Doch wie schafft man ein solches empirisches Fundament? Fundierte empirische Forschung setzt zunächst einmal fundierte Messung voraus, sodass am Anfang der Forschung ein geeignetes Modell zur Nutzenmessung stehen muss.

Man darf hier nicht einfach von der Prämisse ausgehen, der Nutzen von x Gummibären entspräche dem Geldwert, den man zum Erwerb von x Gummibären investieren muss. Eine solche Prämisse würde nämlich die Entscheidungstheorie des erwarteten Werts als gültig voraussetzen. Diese hat sich jedoch für viele

dann aber aufgrund vieler nicht vorhersehbarer Verpflichtungen immer wieder verschoben. Mit knapp dreißigjähriger Verzögerung kann das Versprechen nun doch noch eingelöst werden. Gut Ding will Weile haben!

Bereiche menschlichen Entscheidungsverhaltens als unhaltbar erwiesen (siehe Coombs et al., 1970, Kap. 5).

Ein alternativer Ansatz wurde deshalb von Mosteller und Nogee (1951) vorgeschlagen. Ihr Messmodell geht von der Theorie des erwarteten Nutzens aus, wonach Menschen genau dann indifferent gegenüber zwei möglichen Entscheidungsalternativen sind, wenn beide Alternativen denselben erwarteten Nutzen haben. Der erwartete Nutzen einer Entscheidungsalternative ist dabei die Summe der mit ihrer Eintretenswahrscheinlichkeit p_j gewichteten Nutzenwerte $u(x_j)$ von möglichen Entscheidungskonsequenzen x_j . Nehmen wir beispielsweise an, einer Person wird ein Spiel angeboten, bei dem sie x Gummibären (GB) gewinnen kann, wenn bei einem Würfelwurf ein Ereignis mit der Wahrscheinlichkeit p resultiert (z. B. $p = 1/6$, wenn eine 6 Gewinn bedeutet). Sie muss jedoch 5 Gummibären aus ihrem Vorrat abgeben, wenn der Würfelwurf nicht erfolgreich ist. Der erwartete Nutzen dieses Spiels wäre dann

$$EU(p, x) = p \cdot u(x \cdot GB_{\text{gewonnen}}) + (1 - p) \cdot u(-5 \cdot GB_{\text{verloren}}) \quad (1)$$

Die Person kann entscheiden, ob sie das Spiel annimmt oder nicht, wobei Letzteres dazu führt, dass die Person keinen Gummibären bekommt und auch keinen abgeben muss. Dies hätte den Nutzen u (0 Gummibären). Nach der Theorie des erwarteten Nutzens sollte die Person das Würfelspiel immer dann annehmen, wenn $EU(p, x) > u(0)$, und immer dann ablehnen, wenn $EU(p, x) < u(0)$. Sie sollte genau dann indifferent sein, wenn $EU(p, x) = u(0)$. Wir können also für eine bestimmte Gewinnwahrscheinlichkeit p die Anzahl Gummibären x suchen, bei der die betreffende Person dem Spiel gegenüber indifferent ist, d. h. es in genau 50% der Fälle annimmt und in 50% der Fälle ablehnt. Für diesen Indifferenzpunkt muss nach der Theorie des erwarteten Nutzens gelten (die Einheit GB wird hier zwecks Verbesserung der Lesbarkeit fortgelassen):

$$u(0) = EU(p, x) = p \cdot u(x) + (1 - p) \cdot u(-5) \quad (2)$$

Da wir eine Nutzenmessung auf Intervallskalenniveau anstreben, können wir den Nullpunkt der Nutzenskala willkürlich auf $u(0) = 0$ und ihre Einheit auf $u(-5) = -1$ festlegen. Setzt man diese Annahmen in Gleichung (2) ein, erhält man nach geeigneter Termumformung den Nutzenwert von x Gummibären als Funktion der Gewinnwahrscheinlichkeit p :

$$u(x) = \frac{(1 - p)}{p} \quad (3)$$

Man kann nun das Gummibären-Würfelspiel für verschiedene einfache Würfelereignisse mit unterschiedlichen Gewinnwahrscheinlichkeiten p und assoziierten Indifferenzpunkten x wiederholen, sodass letztlich eine komplette Nutzenfunktion für alle Gewinnwahrscheinlichkeiten im Argumentbereich $1/6 \leq p \leq 5/6$ empirisch abgeschätzt werden kann. Wird eine höhere Auflösungsgenauigkeit gewünscht, ist der einfache Würfelwurf durch ein komplexeres Zufallsereignis zu ersetzen (z. B. den gleichzeitigen Wurf mehrerer Würfel).

Für eine erste Anwendung der skizzierten Methode mag der einfache Würfelwurf ausreichen. Wir wollen mit dieser Anwendung eine fundierte Antwort auf die viel diskutierte Frage geben, ob Gummibären unterschiedlicher Farbe durch unterschiedliche Nutzenfunktionen charakterisiert sind. Die Theorie des erwarteten Werts legt die Hypothese nahe, dass Gummibären unabhängig von ihrer Farbe den gleichen Nutzen haben, weil sie auch dem gleichen monetären Wert entsprechen. Wir werden sehen, ob diese Hypothese auch aus nutzentheoretischer Perspektive empirisch haltbar ist.

Das folgende Experiment wurde nicht präregistriert und es wurde auch keine Poweranalyse zur A-priori-Stichprobenumfangsberechnung durchgeführt. Da die Ergebnisse ohnehin getrennt für jeden Teilnehmer ermittelt werden, um individuelle Gummibären-Nutzenfunktionen zu ermitteln, ist die genaue Anzahl der Versuchsteilnehmer letztlich ohne Belang. Die Frage der Stichprobengröße betrifft nicht die Präzision und Validität der Ergebnisse, sondern lediglich deren Generalisierbarkeit.

Experiment 1

Method

Versuchsteilnehmer: $N = 1$, SF (weiblich), erklärte sich bereit, an dem im folgenden geschilderten Entscheidungsexperiment teilzunehmen. Für die Teilnahme erhielt sie eine Kompensation in Form von Naturalien (360 g Goldbären).

Material: Zwei 360-g-Pakete Gummibären der Marke Goldbären von Haribo wurden für den Versuch verwendet. Die Gummibären des ersten Pakets wurden zunächst nach Farben sortiert (farblos, grün, gelb, orange und rot) und in unterschiedliche Schalen eingefüllt (siehe Abb. 1). Das zweite Paket diente als Reserve für den Fall, dass die Gummibären einer bestimmten Farbe nicht ausreichen. Weiterhin wurden für das Experiment ein Würfel, ein Notizzettel für die Versuchsergebnisse in den einzelnen Durchgängen und ein Stift benötigt.

Design: Unabhängige Variablen waren die Gewinnwahrscheinlichkeit für jedes angebotene Spiel (5 gleichabständige Stufen von $p = 1/6$ bis $p = 5/6$) und die Farbe der potenziell gewinnbaren Gummibären (4 Stufen: grün, gelb, orange und rot). Beide Faktoren wurden in einem vollständig gekreuzten 5×4 -Design innerhalb der Versuchsteilnehmer variiert. Abhängige Variable war der Indifferenzpunkt x für die betreffende Zelle des 5×4 -Versuchsplans, d. h. die Anzahl von Gummibären der betreffenden Farbe, bei der die Versuchsteilnehmer eine Annahmewahrscheinlichkeit von .50 für das angebotene Spiel zeigten.

Prozedur: Der Versuchsaufbau wird in Abbildung 1 illustriert. Nach einer ausführlichen Instruktion in die Regeln des Gummibären-Würfelspiels (wie zuvor geschildert) bekam die Versuchsteilnehmerin alle 5×4 Entscheidungsprobleme (siehe Design-Abschnitt) in vollständig randomisierter Reihenfolge wiederholt dargeboten. Wenn sie ein angebotenes Spiel annahm und beim Würfelspiel verlor, musste sie 5 farblose Gummibären abgeben, von denen sie zuvor ausreichend viele erhalten hatte. Der Nutzen von 5 verlorenen farblosen Gummibären wurde als Einheit der Nutzenskala mit $u(5 \text{ verlorene farblose Gummibären}) = -1$ festgelegt, der Nullpunkt der Skala mit $u(0 \text{ farblose Gummibären}) = 0$. Man beachte, dass farblose Gummibären nur als „einheitliches Zahlungsmittel“ fungierten, selbst aber

nicht gewonnen werden konnten. Dies hat den Vorteil, dass der Nutzen von Gummibären der gewinnbaren Farben auf derselben neutralen Nutzenskala gemessen werden kann.

Abweichend vom klassischen Vorgehen von Mosteller und Nogee (1951) wurden die Indifferenzpunkte auf der psychometrischen Funktion nicht über die zeit- aufwändige Methode konstanter Stimuli entwickelt, sondern über ein modernes, effizientes Staircase-Verfahren, das den Indifferenzpunkt für die betreffende Versuchsbedingung adaptiv ermittelte. Lehnte die Versuchsteilnehmerin ein Spiel ab, wurde x im nächsten Durchgang derselben Versuchsbedingung um 1 erhöht, nahm sie das Spiel an, wurde x um 1 gesenkt (Single-Step-Algorithmus, siehe Meese, 1995). Ein Umkehrpunkt im Staircase-Verfahren liegt vor, wenn auf einen Anstieg



Abbildung 1. Versuchsaufbau bei der Nutzenskalierung von Gummibären

von x ein Abfall folgt (Maximum-Umkehrpunkt) oder wenn umgekehrt auf einen Abfall ein Anstieg folgt (Minimum-Umkehrpunkt). Wenn das Staircase-Verfahren zu zwei derartigen Umkehrpunkten führt (auf ein Maximum folgt ein Minimum oder umgekehrt), wird das Verfahren für die betreffende Versuchsbedingung beendet und der Mittelwert der beiden Umkehrpunkte als Indifferenzpunkt akzeptiert (Meese, 1995, S. 269, Fig. 1). Der gesamte Versuch war beendet, nachdem für alle 5×4 Versuchsbedingungen der betreffende Indifferenzpunkt nach den Regeln des Single-Step-Algorithmus bestimmt war. Wie Meese (1995) gezeigt hat, hat der Single-Step-Algorithmus schon bei Verwendung von nur zwei Umkehrpunkten ausreichende Genauigkeit zur Ermittlung von Indifferenzpunkten.

Ergebnisse und Diskussion

Abbildung 2 zeigt – separat für jede Gummibärenfarbe – die gemäß Gleichung (3) nach der Single-Step-Methode ermittelten Nutzenwerte für die Gummibärenanzahl, bei der die Versuchsteilnehmerin SF jeweils indifferent hinsichtlich der Frage war, ob sie das betreffende Würfelspiel annehmen oder ablehnen sollte. Die fünf Datenpunkte entsprechen dabei den auf der x-Achse abgetragenen beobachteten Indifferenzpunkten zu den fünf Gewinnwahrscheinlichkeiten $p = 5/6$ (ganz links) bis $p = 1/6$ (ganz rechts in der betreffenden Abbildung), jeweils abgestuft in Schritten von $1/6$. Getrennt für jede der vier Gummibärenfarben Grün, Gelb, Orange und Rot wurde außerdem eine Exponentialfunktion der Form

$$u(x) = b_0 + b_1^x \quad (4)$$

an die empirisch ermittelten Nutzenwerte angepasst und in Abbildung 2 jeweils durch eine durchgezogene schwarze Linie veranschaulicht. Zum Vergleich wurde im unteren Teil der Abbildung auch die lineare Nutzenfunktion für farblose Gummibären veranschaulicht, die sich ergeben würde, wenn man farblose Gummibären gewinnen wie auch verlieren könnte. Hier wurde davon ausgegangen, dass die Theorie des erwarteten Werts gilt. Man wäre zwischen zwei Spielen um farblose Gummibären auf der Gewinn- wie auf der Verlustseite also genau dann indiffe-

rent, wenn die erwartete Anzahl gewonnener Gummibären bei beiden möglichen Ausgängen identisch ist.

Ein Blick auf Abbildung 2 bestätigt, dass alle Nutzenfunktionen für farbige Gummibären konvex sind: Der Nutzen steigt mit zunehmender Anzahl von Gummibären überproportional, sodass sich insgesamt ein positiv beschleunigter Kurvenverlauf ergibt. Im Vergleich zu farblosen Gummibären bedeutet das, dass farbige Gummibären in geringer Anzahl (bis ca. 5 Gummibären) zwar einen etwas geringeren Nutzen als farblose Gummibären haben, aber mit zunehmender Anzahl deutlich höhere Nutzenwerte erzielen als die gleiche Anzahl farbloser Gummibären.

Wie lässt sich dieses Phänomen der „Nutzenexplosion“ farbiger Gummibären mit zunehmender Kardinalität der Gummibärenmenge erklären? Eine naheliegende Erklärung, die zunächst auch durch Selbstberichte der Versuchsteilnehmerin gestützt wurde, ist der Geschmack der Gummibären. Möglicherweise schmecken farbige Gummibären einfach besser als farblose Gummibären. Wenn dem so wäre, dann sollten farbige und farblose Gummibären allerdings bei Blindverköstigung klar unterscheidbar sein. Um dies zu überprüfen, führten wir ein weiteres Experiment durch.

Experiment 2

Method

Versuchsteilnehmer: $N = 1$, SF (weiblich), wie in Experiment 1. Für die Teilnahme erhielt SF erneut eine Kompensation in Form von Naturalien (150 g saure Goldbären).

Material: Ein 360-g-Paket Gummibären der Marke Goldbären wie in Experiment 1; die Gummibären wurden in Experiment 2 allerdings nicht nach Farben sortiert, sondern in einem gemeinsamen Behälter aufbewahrt wurden.

Design: Es wurde ein einfaktorielles Innersubjekt-Design mit dem 5-stufigen Faktor Gummibärenfarbe (farblos, grün, gelb, orange und rot) verwendet. Abhängige Variable war das (verblindete) Farbidentifikationsurteil mit den möglichen Urteilen „farblos“, „grün“, „gelb“, „orange“ und „rot“.

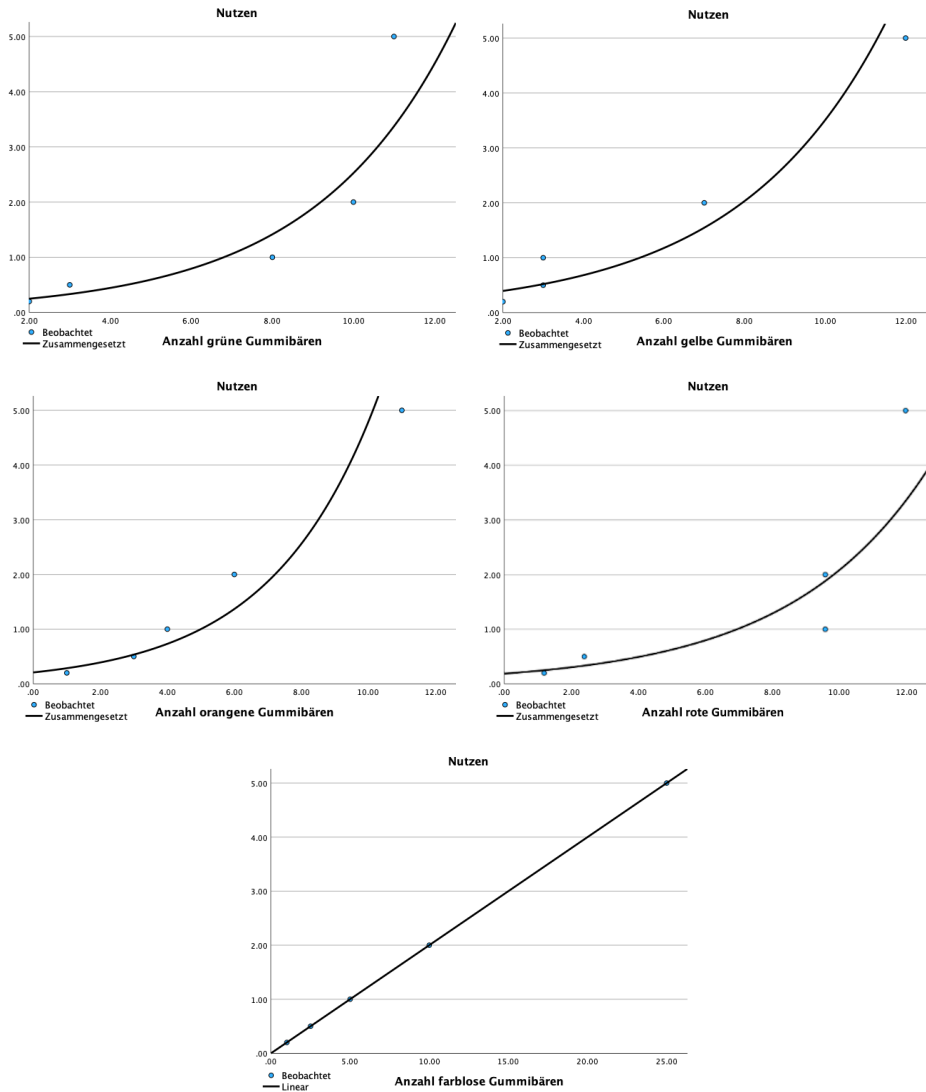


Abbildung 2. Nutzenfunktionen von Gummibären unterschiedlicher Farben (obere 4 Diagramme) und die Referenznutzenfunktion für farblose Gummibären (unteres Diagramm)

Prozedur: Die Versuchsteilnehmerin hatte zunächst die Möglichkeit, den Behälter mit den Gummibären einzusehen und sich somit mit der Farbzusammensetzung und -häufigkeit der Stimuli vertraut zu machen. Ihr wurden sodann mithilfe eines Schals die Augen verbunden. Es wurde geprüft und validiert, dass sie durch diesen Schal nichts sehen konnte. Danach wurden durch Zufallsziehung aus dem Behälter insgesamt $M = 60$ Gummibären gezogen (ohne Zurücklegen!) und SF sequenziell zur Geschmacksprobe dargereicht. SF gab sodann pro Gummibär eines der o. g. fünf Farbidentifikationsurteile ab. Sie entschied selbst, wann der nächste Gummibär zu verabreichen ist. Die künstliche Verblindung mittels Schal wurde nicht zwischen zwei Stimuli, sondern erst im Anschluss an eine komplette Urteilssitzung

Tabelle 1. Konfusionsmatrix zu Farbidentifikationsurteilen bei Blindverköstigung von Gummibären.

Dargebotener Gummibär	Verblindetes Farbidentifikationsurteil					Summe
	farblos	grün	gelb	orange	rot	
farblos	1 (10%)	4 (40%)	3 (30%)	1 (10%)	1 (10%)	10
grün	2 (25%)	1 (12.5%)	0 (0%)	0 (0%)	5 (50%)	8
gelb	1 (14.3%)	2 (28.6%)	0 (0%)	2 (28.6%)	2 (28.6%)	7
orange	2 (14.3%)	1 (7.1%)	2 (14.3%)	1 (7.1%)	8 (57.1%)	14
rot	5 (23.8%)	3 (14.3%)	1 (4.8%)	4 (19%)	8 (38%)	21

Anmerkung. Angegeben sind die Häufigkeiten (und Zeilenprozent) von Farbidentifikationsurteilen sowie die Zeilensummen von Urteilen (in der letzten Spalte). Korrekte Urteile in der Hauptdiagonalen der Matrix sind fett markiert. Als weitere Erklärung für die konvexen Nutzenfunktionen wäre ein visuell-ästhetischer Vorteil farbiger gegenüber farblosen Gummibären denkbar. Wenn dem so ist, dann sollte das abschließende Schönheitsrating zwischen farbigen und farblosen Gummibären diskriminieren. Dies ist tatsächlich der Fall: SF vergab für rote und orangene Gummibären das Rating 5, für gelbe und grüne das Rating 3 und für farblose das Rating 2. Auch wenn sich diese Unterschiede bei nur einem Likert-Rating pro Gummibärenfarbe nicht auf Signifikanz prüfen lassen, ist doch offenkundig, dass farbige Gummibären höhere Schönheitsratings erzielen als farblose. Auch innerhalb der Gruppe farbiger Gummibären ergibt sich für die Gummibären mit den höchsten Schönheitsratings (rot, orange) die ausgeprägteste positive Beschleunigung der Nutzenfunktion.

aufgehoben. Das Experiment wurde in 6 verschiedenen Sitzungen mit je 10 Gummibären verteilt über 4 Tage durchgeführt. Unmittelbar im Anschluss an die sechste Sitzung wurde als Kontrollvariable ein Rating der Schönheit eines exemplarischen Gummibären jeder der fünf verwendeten Farben erhoben (5-stufige Likert-Skala von 1 = *hsslich* bis 5 = *hbsch*).

Ergebnisse und Diskussion

Insgesamt waren nur 11 der 60 Farbidentifikationsurteile (d. h. 18%) korrekt, was knapp unter der Zufallstrefferwahrscheinlichkeit von 20% bei fünf möglichen Farburteilen liegt. Auch wenn man die Trefferwahrscheinlichkeiten der einzelnen Farben separat betrachtet, ergibt sich für 4 der 5 Farben eine Trefferwahrscheinlichkeit unter dem Zufallsniveau. Lediglich rote Gummibären wurden in immerhin 38% der Fälle korrekt beurteilt, was zwar knapp signifikant über dem Zufallsniveau liegt (upper-tail $p = .043$ gemäß Binomialtest für $p = .20$ bei $N = 21$), nach Bonferroni-Adjustierung der $\alpha = 0.05$ -Fehlerwahrscheinlichkeit ($\alpha^* = \alpha/5 = 0.01$) aber ebenfalls als insignifikant zu bewerten ist.

Tabelle 1 liefert eine Zusammenfassung aller Ergebnisse in Form einer Konfusionsmatrix, bei der die dargebotenen Stimuli in den Zeilen und die Farburteile in den Spalten abgetragen werden. Wie die Zeilensummen klar zeigen, hat die Versuchsteilnehmerin die ungleichen Basisraten der verschiedenen Farben klar erfasst, was möglicherweise auch ursächlich dafür ist, dass nahezu alle korrekten Urteile auf die am häufigsten dargebotene Farbe Rot entfallen. Dennoch ergibt sich keine Evidenz für Farbdiskriminationsfähigkeit von SF bei verblindeter Verköstigung. Die Hypothese, dass der konvexe Verlauf der Nutzenfunktion für farbige Gummibären auf einen reinen Geschmacksvorteil zurückzuführen ist, der von visueller Information unabhängig ist, muss also eindeutig verworfen werden.

Gesamtdiskussion

Unsere Experimente zeigen, dass (a) farbige Gummibären relativ zu farblosen durch positiv-beschleunigte Nutzenfunktionen charakterisiert sind und (b) der

Grad der positiven Beschleunigung mit der Beurteilung der Schönheit des Gummibären korreliert. Zusammengefasst mit dem Befund von Experiment 2, dass Gummibären unterschiedlicher Farbe bei Blindverköstigung geschmacklich nicht unterscheidbar sind, stützt dies die Theorie, dass die „Nutzenexplosion“ von farbigen Gummibären mit zunehmender Anzahl auf ihrer visuell wahrgenommenen Schönheit und nicht auf ihrem Geschmack beruht. Wird die wahrgenommene Schönheit auf den Geschmack des betreffenden Gummibären attribuiert – eine Tendenz, die ursprünglich auch bei der Versuchsperson SF festzustellen war –, so handelt es sich daher eindeutig um eine Fehlattribution – mithin also um eine Geschmacksillusion. Wir halten fest: Neben der Anzahl hat allein die Farbe einer Menge von Gummibären maßgeblichen Einfluss auf ihren Nutzen.

Limitationen

Trotz des klaren Ergebnismusters dürfen die Limitationen der berichteten Experimente nicht in intransparente Tüten verpackt werden. So wurden auf der versuchsplanerischen Seite zwar geschmackliche und visuelle Aspekte von Gummibären sorgfältig kontrolliert. Akustische Aspekte blieben aber völlig unkontrolliert. Zwar berichteten weder SF noch der Versuchsleiter EE von Geräuschempfindungen, die durch Gummibären ausgelöst wurden und möglicherweise sogar mit ihrer Farbe konfundiert sein könnten. Aber das besagt insofern nicht viel, als eine Stimulierung im Infra- oder Ultraschallbereich nicht ausgeschlossen werden kann, die zwar für menschliche Ohren nicht hörbar ist, aber dennoch die Nutzenurteile in subtiler Weise beeinflussen kann. Dass diese Idee nicht völlig von der Hand zu weisen ist, zeigt die bereits von Hoepfner (1997) vorgelegte Evidenz für die Kommunikationsfähigkeit von Gummibären, die lange Zeit fälschlich ins Reich der Fantasie verwiesen wurde, weil sie für menschliche Ohren nicht hörbar ist.

Weiterhin muss beachtet werden, dass die von uns verwendete Nutzenskalierung die Theorie des erwarteten Nutzens voraussetzt, die mit einer Reihe von Befunden zu Entscheidungen unter Unsicherheit nicht im Einklang steht. Auch wenn viele Befunde nahelegen, dass subjektive Wahrscheinlichkeiten (Probability Weights)

nicht identisch mit objektiven Wahrscheinlichkeiten sind (Kahneman & Tversky, 1979), so denken wir nicht, dass der festgestellte konvexe Verlauf der Nutzenfunktionen durch Abweichungen von einer Identitätsfunktion zwischen subjektiven und objektiven Wahrscheinlichkeiten erklärbar ist. Dies gilt zumindest für den hier interessierenden Bereich zwischen $p = 1/6$ und $p = 5/6$ (Gonzalez & Wu, 1999).

Eine weitere Limitation ist die eingeschränkte Generalisierbarkeit der Ergebnisse, die sich aus der geringen Zahl von nur einer bislang untersuchten Versuchsteilnehmerin ergibt. Replikationsexperimente sind unbedingt erforderlich. Sie sollten möglichst mit Probanden aus verschiedenen Staaten, Kulturen und Subkulturen durchgeführt werden, um dem ausgeprägten WEIRD-Bias² in der empirischen psychologischen Forschung entgegenzuwirken.

Schließlich dürfen wir nicht übersehen, dass der hier verwendete Forschungsansatz Gummibären lediglich als Objekte, nicht aber als Subjekte von Entscheidungen begreift. Wie würde beispielsweise die Nutzenfunktion für rote Gummibären aus der Sicht eines farblosen Gummibären aussehen? So wichtig diese Frage ist, so wenig können wir mangels geeigneter Daten gegenwärtig zu ihrer Beantwortung beitragen. Die Fortschritte der Forschung zur Kommunikation von Gummibären (Hoepfner, 1997) lassen jedoch die Hoffnung nicht unbegründet erscheinen, dass wir eines Tages empirisch begründete Entscheidungstheorien vorlegen können, die Gummibären nicht auf Objekte psychologischer Forschung reduzieren, sondern sie als Subjekte von Entscheidungen begreifen und untersuchen.

Fazit

Nutzenfunktionen für farbige Gummibären – insbesondere für rote und orangefarbene Gummibären – sind konvex und können gut durch Exponentialfunktionen beschrieben werden. Dieses zentrale Ergebnis unserer Untersuchungen hat erhebliche Implikationen nicht nur für die Grundlagenforschung zu Gummibären, sondern

² WEIRD = western (westlich), educated (gebildet), industrialized (industrialisiert), rich (wohlhabend, reich) und democratic (demokratisch) (<https://de.wikipedia.org/wiki/WEIRD>).

auch für die angewandte Forschung. Dies gilt beispielsweise für die Konsumentenpsychologie. So ist nach unseren Befunden zu erwarten, dass Konsumenten eine Tüte mit ausschließlich farbigen Gummibären im Vergleich zu einer Tüte mit farblosen Gummibären als nützlicher bewerten. Dieser Effekt kann noch dadurch gesteigert werden, dass man in der Tüte farbiger Gummibären die roten und orangefarbenen überrepräsentiert. Genau dies scheint in der Praxis auch zu passieren, wie ein Blick auf Abbildung 1 offenbart. Da die Kalorienanzahl von Gummibären nur von ihrer Menge und nicht von ihrer Farbe abhängt, ist die Zusammenstellung von Angeboten mit vielen roten und orangefarbenen Gummibären ein einfaches und kostengünstiges Mittel, um den Nutzen von Gummibären für eine gegebene Kalorienanzahl zu maximieren. Diese Einsicht schont nicht nur die Finanzen des Konsumenten und des Produzenten, sondern ist letztlich auch der Volksgesundheit zuträglich.

In jedem Fall stützen unsere Forschungsergebnisse die Hypothese, dass in Folie verpackte Schokoladenbären kein Konkurrenzprodukt sind, welches Umsatzeinbußen auf dem Gummibärenmarkt mit sich bringen könnte. Dies gilt umso mehr, wenn Schokoladenbären in goldene Folie statt in rote Folie verpackt werden. Die Kosten für den Rechtsstreit vor dem Bundesgerichtshof im Jahr 2015 (Radomsky, 2015) hätten also vermieden werden können, wenn man einen kleinen Teil dieser Finanzmittel im Vorfeld in entscheidungspsychologische Grundlagenforschung zu Gummibären investiert hätte.

Postscript

Jochen Musch und Edgar Erdfelder schließen nicht aus, dass der vorliegende Text zur Gänze oder in Teilen mithilfe von GBT-4 (der derzeit aktuellsten Version des Gummi-Bären-Transformers) verfasst wurde. Entsprechend beanspruchen sie nicht die alleinige Autorenschaft, können aber auch nicht die Verantwortung für die Korrektheit aller Literaturangaben übernehmen.

Referenzen

- Bouchard, T. (1989). *The Minnesota study of bears reared apart*. Janus.
- Coombs, C. H., Dawes, R. M., & Tversky, A. (1970). *Mathematical psychology*. Prentice-Hall.
- Chomsky, N. (2023, 1. April). Mouths without a brain. On the ridiculousness of the recent GBT hype. *Washington Post*, S. 86.
- Döring, N. (1997). Die sexuellen Phantasien der Gummibärchen. *Sexological Review*, 31, 11–29. Bonn University. Psychological Institute. Internet Archive Wayback Machine. Abgerufen am 12.5.2023. <https://web.archive.org/web/19980708122811/http://www.psychologie.uni-bonn.de/sonstige/gummibaer/fantasie.htm>
- Dörner, D. (1989). *Die Logik des Misslingens*. Rowohlt.
- Endepohls, M., & Funke, J. (2004). *Psychotherapie bei Gummibären*. Ruprecht-Karls-Universität. Psychologisches Institut. Abgerufen am 14.5.2023 auf <https://www.psychologie.uni-heidelberg.de/ae/allg/forschun/gb/therapie.html>
- Funke, J. (1995). *Komplexes Problemlösen bei Gummibären*. Ruprecht-Karls-Universität. Psychologisches Institut. Abgerufen am 16.5.2023. <https://www.psychologie.uni-heidelberg.de/ae/allg/forschun/gb/komplex.html>
- Funke, J., & Gerdes, H. (1995). *Gummibären-Forschung.de*. Internet Archive Wayback Machine. Abgerufen am 10.5.2023 unter <https://web.archive.org/web/20160304091025/http://www.gummibaeren-forschung.de/>
- Gonzalez, R., & Wu, G. (1999). On the shape of the probability weighting function. *Cognitive Psychology*, 38, 129–166. <https://doi.org/10.1006/cogp.1998.0710>
- Hawking, S. (2020). On the use of quantum gum for the creation of interstellar wormholes. *Science*, 130, 14–20.
- Hoeppepner, W. (1997). *Und sie kommunizieren doch: Sprachliche Varietäten bei Gummibären*. TU Dortmund. Journal of Applied Idiosyncrasies. Abgerufen am 10. 3. 2023. <https://www-ai.cs.tu-dortmund.de/FUNSTUFF/JAI/gummibaer.html>
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–291. <https://doi.org/10.1017/CBO9780511609220.014>

- Knoke, M. (2009, 9. April). *Das Sexualleben der Gummibären*. Der Spiegel. Abgerufen auf Spiegel Online am 12.05.2023, <https://www.spiegel.de/lebenundlernen/uni/ju-x-forschung-das-sexualleben-der-gummibaeren-a-618218.html>
- Meese, T. S. (1995). Using the standard staircase to measure the point of subjective equality: A guide based on computer simulations. *Perception & Psychophysics*, 57(3), 267–281. <https://doi.org/10.3758/BF03213053>
- Miller, G. (1995). The magical number five, plus or minus two. *Psychological Review*, 45, 344–350.
- Mommsen, H. (1995). *Geschichte der Gummibären im 20. Jahrhundert*. Herder.
- Mosteller, F., & Nogee, P. (1951). An experimental measurement of utility. *Journal of Political Economy*, 59, 371–404. <https://doi.org/10.1086/257106>
- Musch, J. (1997). *Die ersten zwanzig Jahre: Eine selektive Zusammenfassung ausgewählter Befunde der Gummibärchenforschung*. Gummibären-Forschung.de. Abgerufen am 12.5.2023 auf <https://web.archive.org/web/20011012164700/http://www.gummibaeren-forschung.de/musch/musch.htm>
- Nosek, B. (2015). Recommendations for increasing replicability in gummibear research. *European Journal of Personality*, 17, 1–14.
- Tavladaki, T., Fitrolaki, M.-D., Spanaki, A.-M., Ilia, S., Geromarkaki, E., & Briassoulis, G. (2012). The sweet lung: Chewing gummi bear aspiration. *Lung India*, 29(3), 267–269. <https://doi.org/10.4103/0970-2113.99113>
- Throckmorton, G. S., Ellis, E., & Hayasaki, H. (2004). Masticatory motion after surgical or nonsurgical treatment for unilateral fractures of the mandibular condylar process. *Journal of Oral and Maxillofacial Surgery*, 62, 127–138. <https://doi.org/10.1016/j.joms.2003.01.003>
- Radomsky, S. (2015, 23. September). *Treffen sich zwei Bären vor Gericht*. Süddeutsche Zeitung. Abgerufen am 10. 3. 2023. <https://www.sueddeutsche.de/geld/bgh-urteil-treffen-sich-zwei-baeren-vor-gericht-1.2660710>
- Walter, U. (2016, 30. März). *Weshalb Sie sich den 2. August 2048 notieren sollten*. WELT. Abgerufen auf WELT Online am 5.5.2023. <https://www.welt.de/wissenschaft/artikel160310635/Weshalb-Sie-sich-den-2-August-2048-notieren-sollten.html>

The Role of Credulity in Failed Problem Solving

ROBERT J. STERNBERG

Cornell University New York, USA

Abstract. Credulity is the predisposition to be easily, and often too easily, accepting of new information and arguments in the absence of evidence that the information is, or the arguments are, credible. I suggest that the greatest problem faced today by many problem solvers is not lack of problem solving knowledge or skills but, rather, credulity. Authoritarian leaders know this and take advantage of it. IQs rose 30 points during the 20th century, but despite this, credulity is at astonishingly high levels. If people are credulous, then all their problem solving skills can be for naught. I discuss the problem and suggest solutions to it.

Because I am a university professor, most of my friends and colleagues are also university professors. A few are in allied fields, such as academic administration and elementary/secondary education. And then there is Joachim Funke, who spent most of his life as a university professor but, improbably, is also a DJ (disc jockey). As an adolescent, I had some favorite DJs—Cousin Brucie and Harry Harrison on WABC Radio, New York City—but I never expected actually to meet one. And then along came Jofu, as his friends call him (and I hope I am included in their number).

Complex Problem Solving: The European Tradition

Jofu has made many different contributions to psychology, but I will focus on one—his research on complex problem solving. Once upon a time, there were two major approaches to research on complex problem solving. The “American” perspective emphasized fairly well-structured problems, such as the Missionaries and Cannibals problem and the Tower of Hanoi problem (Sternberg & Frensch, 1991). The approach was probably best represented by the classic work of Newell and Simon (1972) on human problem solving.

The American approach, I believe, declined because it had no viable future. The problems were too unlike real-world problems. They suffered from many of the same issues as conventional intelligence tests (Sternberg, 2020b). They were well-structured, decontextualized, somewhat sterile, unrealistic, emotionally uninvolved, for low or no stakes, simplistic, and characterized by unique or almost unique paths to solution. In contrast, problems studied in the European, and mostly German, tradition were richer and realistic. They resembled problems in the world whose solutions matter.

It is always hard to pin the start of a tradition to any one person’s work, but certainly, the work on Lohhausen initiated by Dietrich Dörner and his colleagues would be considered seminal (e.g., Bick et al., 1994). Being mayor of a city, as in the work on the Lohhausen problem, more reflects actual life challenges or organizations and competing interests than does the work on the Missionaries and Cannibals problem.

As a professor in an American university, I am scarcely in a position to assess the state of German work on problem solving, but from my vantage point on the other West-to-East side of the Atlantic Ocean, Jofu would appear today to be the most eminent, or certainly among the handful of most eminent, immediate successors to Dörner. His work has investigated, in a somewhat more structured way than his predecessors, the psychological processes involved in complex problem solving. It has been a great success, as shown by the roughly 10,000 citations of his work on Google Scholar.

Much of my own work in recent years has been at the very least related to work in the complex European tradition (e.g., Sternberg, 2020a; Sternberg & Sternberg, 2017; Sternberg et al., 2017, 2019, 2021). My colleagues and I have examined complex problem solving in the contexts of scientific problem solving and intercultural problem solving. I have also looked at real-world problem solving as the basis for what I have come to call adaptive intelligence (Sternberg, 2021a, 2022). I further have suggested that two processes—seeking internal coherence and seeking external correspondence—underlie adaptive intelligence (Sternberg, 2021b). Internal coherence is the property of the parts of the whole of a solution fitting together logically or empirically. External correspondence is the property of the solution fitting the facts of the real world.

My goal in this particular essay is not to review my own work on problem solving but, rather, to discuss a fundamental obstacle to the kinds of cognitive approaches to problem solving many of us, including myself, have taken. I question whether, when I write in mid-2022, the greatest challenge of studying complex problem solving is as much studying people's ability to intelligently solve complex problems as it is studying their desire to do so.

Laying Out the Welcome Mat for Authoritarians

We are living in a time when authoritarians are making a comeback—not only in far-flung places such as China, Nicaragua, Venezuela, Iran, and Saudi Arabia but also in Europe: in Hungary, Poland, and most of all, Russia, which has descended from being merely authoritarian to being a brutal dictatorship reminiscent of Axis countries during World War II. It may be a coincidence or it may be that something more is going on. Young, and even middle-aged and most older people today did not live through World War II. To them, authoritarian leaders of the past are, well, past. Democracy presents its own set of challenges. What if we are at a point in a cycle where people know the challenges of democracy and have tired of democracy and its challenges but do not well know the challenges of authoritarianism? What is for sure is that democracy is on the decline (Albright, 2018; Coppedge et al.,

2022; Klass, 2017; Levitsky & Ziblatt, 2018; Mounk, 2018). And the authoritarians are getting elected, not elevating themselves to power through coups d'état. They may stay in power through attempts to falsify election results, as Donald Trump tried to do (Chowdhury et al., 2022) and as Jair Bolsonaro in Brazil appears to be attempting to prepare to do (Ecarma, 2022). But initially, at least, they often get themselves elected through popular will.

What if, as seems to be happening, the problems of the world seem overwhelming and it becomes much easier simply to accept the solutions of others, particularly people one respects—perhaps someone who sets himself up (and it is almost always a “him”) as a populist who will serve the people? Many people are hurting, and they are ready to blame others for their misfortune. They are also ready to hear a false populist claim that he knows the enemy, and the enemy is right there undermining you!

But what then, if the authorities one respects do not think in a respectable way? What if the authorities are people who are not, well, even respectable as people? Then the greatest problem may become one of not falling into the trap of credulity. That, I believe, is the challenge of the present age: credulity. Things have reached such a low point, at least in the United States, that a government official from Arizona, Rusty Bowers, testified about how he was illegally pressured to manipulate the tallying of votes to support Donald Trump in the 2020 US presidential election, and then, after describing the illegal pressure, said he would vote for Trump again if Trump ran again in 2024 (Graham, 2022). Really, how bizarre is that?

Credulity as a Basis for A-rational Thought

Credulity is “a tendency to be too ready to believe that something is real or true” ([Online search engine result for the query “credulity”], n.d.b). It is similar but not quite identical to gullibility. In particular, the difference between credulity and gullibility is that “credulity is a willingness to believe in someone or something in the absence of reasonable proof,while gullibility is the quality of readily

believing information, truthful or otherwise, usually to an absurd extent” (Credulity vs Gullibility—What’s the difference?, n.d.a). Credulity is also different from persuasibility. The question is not how easily you are persuaded but, rather, how easily you are persuaded in the absence or near-absence of evidence, or in the presence of weak or flawed evidence.

Credulity is a combination of a skill and an attitude. One needs the skills to recognize that one is too easily accepting what one is told or what one reads; one needs the attitude of wanting to apply the skill. These skills probably derive from some combination of general intelligence, adaptive or practical intelligence, and rationality. The attitude is largely a desire to hold oneself to a higher standard of thinking than many people are willing to hold themselves. I suggest in this essay that the greater problem is the attitude, not the skill set. For example, despite revelations that Fox News hosts in the United States lied through their teeth and said what they believed their audience wanted to hear, knowing it was false (Friedland, 2023), Fox News is the top cable news channel (Watson, 2023) in terms of viewership. Put another way, many people want to be lied to, so long as the lies confirm what they want to believe. They do not want to disconfirm their false beliefs. Ideology trumps rational thinking. For many people, it is not even a contest. The problem, then, is that often people lie because they realize their audience wants to be lied to (Serwer, 2023).

What is the cause of people’s credulity? There probably is no one cause, but people seem to have a very strong need to belong or affiliate with others who are like-minded (Murray, 1938) and to feel as though they are part of something greater. Credulity meets those needs.

Gilbert (1991) proposed that people, when confronted with new information, go through two basic steps. The first step is to encode and potentially store the information. At this point, one accepts the information as true. The second step is to evaluate the credibility of the information. Someone could be an excellent learner, with a vast store of information and, presumably, a high level of crystallized intelligence (Carroll, 1993; Cattell, 1971) because they are highly adept at Gilbert’s first step; but having high crystallized intelligence does not guarantee that they will

seriously engage the second step of evaluating the information. Even having high fluid intelligence does not guarantee they will do so. Their high fluid intelligence may facilitate the quality of their analysis, but first, the people have to decide that they even want to do the evaluation, and to do it objectively. In the United States, at least, it often is very difficult, and may even be considered somewhat blasphemous, to question orthodoxies. Thus, someone could be highly intelligent with respect to abilities but not with respect to attitudes (Sternberg, 2021a)—they have the abilities but not the disposition to apply those abilities, even sometimes in crucial circumstances. For some, whatever dispositions they had to question orthodoxies were drilled out of them—often in the name of patriotism or religion or ideology—in the home, the school, or society in general. In the final section of this article, I discuss how the disposition to question can be drilled back in.

Fiedler (2012, 2019) has dealt with a related problem in his work on metacognitive myopia. He found that people are accurate in using information but are relatively oblivious to the source of the information. So, if people only read and listen to truth-challenged, flawed media, they may commend themselves for using that information accurately while spouting off false and even ridiculous statements because their sources are factually challenged or, as it turns out in the case of Fox News, purposely lying to them to increase appeal to audience biases and prejudices (Friedland, 2023). Much of the political dysfunction seen today in the world, certainly in the United States, can be attributed to this kind of cognitive myopia and the credulity that results from it. Sadly, people often seek only news outlets that appeal to their already-existing confirmation biases, and so effectively shield themselves from being able to see things in any but their preferred myopic way.

Even having a high level of ability in abstract analysis and the desire to apply it does not always help in discerning what is true. In our own series of studies on scientific reasoning (Sternberg & Sternberg, 2017; Sternberg et al., 2017, 2019, 2020), we found that students' analytical skills, as measured by fluid-ability tests and standardized tests used for college admissions in the US (SAT and ACT), were weak and sometimes even negative predictors of the students' successful

application of these skills as applied to STEM (science-technology-engineering-mathematics) reasoning problems. In other words, IQ-related skills did not hold up when they were applied outside a relatively abstract-analytical domain.

Why Is Credulity Important to Understand?

Why is credulity so important a construct, at least as we approach the middle of the 21st century? There are several major reasons.

First, skillful problem solving (of the kind studied by Joachim Funke, among others) requires an Enlightenment mindset—that the truth matters. If one does not care what the true solution or, at least, an optimal solution to a problem is, then one's problem solving is likely to be defective, no matter what one's level of problem solving skills. In other words, one's attitude toward problem solving may matter at least as much as one's skills. People who are satisfied with Big Lies, as is much of the United States—perhaps 30 % (Devega, 2021), and likely more if one includes big lies on the left as well as the right—are going to be bad problem solvers, no matter what their cognitive skills may be. In Russia, Putin's popularity rating is over 80% while he is responsible for a massive genocide rivaling those of World War II (Statista Research Department, 2022). For them, cognitively-based research on problem solving is largely irrelevant. Cognitive research fails to capture their failure in problem solving. If 80% of a population can approve of genocide, does it matter how they would perform either on the Missionaries and Cannibals problem or on a problem in which they are mayor of a small city? Maybe not, unless one is interested in how they decide which innocent people in their city to execute, based on a fantasy about the town's history. If this sounds too cynical, well, so is genocide being committed with people's approval. One could argue, of course, that they do not know what is happening, but willful ignorance is perhaps not such a great excuse for ignorance (see Goldhagen, 1997). When all news is cut off except the government's line, one can be pretty sure one is being fed propaganda, as anyone who lives in a formerly Communist country should know.

The Russians use a “firehose of falsehood” model for their propaganda (Paul & Mathews, 2016). The main characteristics of this model, according to Paul and Mathews, are that (a) there are many channels of government propaganda, (b) the government is shamefully willing to disseminate outright falsehoods or, at best, partial truths, (c) the information is fed in a rapid, continuous, and repetitive way, and (d) there is no attempt to be consistent. The story can change quickly, as when Russia recently first said it had no intention of invading Ukraine and then invaded anyway. As late as February 21, 2022, the BBC was suggesting that Russia might well not invade (Gardner, 2022). Other news outlets carried similar optimistic analyses. Russia invaded on February 24, three days later.

Second, although propagandists have become better at what they do, it is not at all clear that ordinary laypersons—and probably many professionals in various fields—have become better at mentally defending against propaganda, including outright falsehoods. Indeed, people are often extremely gullible (Forgas & Baumeister, 2019). To the extent that important problems need veridical solutions, people must be able to defend themselves against propaganda, but are often unable to do so (Paul & Mathews, 2016).

Third, people overestimate their ability to discern truth from falsehood (Bond & DePaulo, 2006). Often, the least able and the most credulous have the highest opinion of their skills, the so-called Dunning–Kruger effect (Kruger & Dunning, 1999; Zimmerman, 2016). The point of the effect is that people who are not competent at a job often do not even have the competence to recognize their own incompetence. The result is that people may believe that they are immune to propaganda when, in fact, they are not.

Fourth, intelligence, as defined traditionally in terms of IQ, general intelligence (g), or whatever, is not only not protective against persuasion attempts but can actually be detrimental. The reason is that people who are intelligent believe that intelligence is protective against crude persuasion attempts, whereas it is not. Credulity stems from foolishness, not from low intelligence (Sternberg, 2018). Foolishness is lack of wisdom, not lack of intelligence. Many supporters of Trump’s Big Lie had graduated from top Ivy League universities in the United States. Their

degrees did not save them from foolishness. IQs rose 30 points during the 20th century (Flynn, 1987, 2012, 2016), but these increases in IQ seem to have done little or nothing to reduce people's credulity.

Fifth and finally, toxic leaders count on people's credulity. They invent patently false memes—that the brutal, genocidal invasion of Ukraine is a “special military operation”; that the US presidential election of 2020 was “stolen”; that the way to stop gun violence is with more guns—and millions of people will believe them. The US Supreme Court, on the day I am writing, June 23, 2022, just declared a New York State law restricting guns to be unconstitutional—this after countless mass shootings in the United States that sets it apart from any other country in the world. The number of deaths seems never to be quite high enough for the five Ivy-League-educated Supreme Court justices who voted in favor of more guns. Of course, one who voted for more guns did not go to an Ivy League School but graduated with highest honors from Notre Dame, a prestigious if not Ivy League law school.

The role of toxic leaders in preying upon people's credulity can hardly be overestimated. People are attracted to toxic leaders—to their self-confidence, their charisma, their sense of self-importance, their pretense of caring for people who believe they have been victimized, their seeming infallibility. And toxic leaders know that, to achieve power and stay in power, they need to generate, as much as they can, blind conformity by appealing to the credulity of their followers and of those whom they would like to persuade, or if necessary, force to be followers (Sternberg, 2022, 2023).

Toxic leaders show characteristics personally and in their leadership styles of the dark triad (Paulus & Williams, 2002). They are narcissistic and care only about themselves and their self-aggrandizement. They have little or no empathy or compassion. Life, for them, is a zero-sum game that they are destined to win at everyone else's expense. They also are Machiavellian. They are schemers. They are willing to use almost any means to get their way, no matter how underhanded or amoral. They see other people as tools to exploit for their own purposes. Finally, they are psychopathic. They have no conscience, no concern for others,

no compassion, no remorse. If they fool others, they are happy because it shows what they already know: that they are omnipotent, and others are like clay in their hands. Thus, these leaders are human but show few of the traits that we associate with humanity. Rather, they try to pretend that they care about humanity and often do so successfully because any deception is, to them, fair game. How do dark-triad toxic leaders prey upon people's credulity?

The Mechanisms for Preying on Credulity

There are many possible mechanisms for exploiting people's credulity and thereby bypassing their normal problem solving processes. This list, therefore, is only partial.

Redefining the problem: In recent reports, various Fox News commentators in the United States have claimed that the January 6 hearings by the House Select Committee to Investigate the January 6th Attack on the US Capitol have not "moved the needle" (see Milbank, 2022), meaning, presumably, that they have not changed the situation noticeably. Reports in other news outlets that do not adhere to the "Trump line," such as the New York Times, Washington Post, Boston Globe, and CNN, give a completely different impression.

The question that underlies these columns is: what needle? The most damaging revelations about a president in the history of the United States have not "moved the needle"? Revelations that a US president planned a coup d'état have not "moved the needle"? One needle the revelations may not have moved is the faith of the followers of Donald Trump that they will vote for him if he runs for US president again. That, presumably, was what the Fox pundits were talking about. But the hearings were not about moving poll numbers or about voting plans. They were about a president who openly was planning a coup.

A way to take advantage of credulity is to redefine the problem. Instead of talking about a likely treasonous president, the pundits, seeking to influence their listeners to welcome an authoritarian, talked about support of Trump's devotees or of Republicans more generally, many or most of whom are devotees. Pinochet and

his devotees talked about his economic success, not his authoritarian government and brutal suppression of dissent (Codevilla, 1993). Putin talks about destiny and Peter the Great, not about genocide (MOCKBA, 2022). President Joe Biden talked about how the drawdown of US troops from Afghanistan proceeded “in a secure and orderly way” (The White House, 2021), which could not have been further from the truth. The withdrawal was marked by utter chaos and disorder (Shepp, 2021) and left the country bereft and subject to a brutal takeover by the Taliban. In the fashion of a true politician who blames the last guy, Biden’s secretary of state, Anthony Blinken, blamed Biden’s predecessor, Donald Trump, for the incompetence of Biden’s withdrawal plan (Desiderio, 2021). So, a disaster of Biden’s making was redefined as an orderly transition where any problems were due not to Biden but to Trump.

It would be a great lesson for young people if any politician who made a mistake would actually own up to the mistake without being forced to do so, as Boris Johnson, Prime Minister of the United Kingdom, was forced to for parties held at 10 Downing Street during lockdown (Diffley, 2022). Johnson also, predictably, announced he would not resign. Young people might learn that everyone makes mistakes and that, rather than redefining the blame, some adults actually take it upon themselves without being forced to do so.

Ideological resonance: When it comes to ideology, IQ doesn’t matter. The judges of the US Supreme Court almost all, or possibly all, have very high IQs. It is not trivial to be appointed to the highest court in the nation. The large majority, as noted above, have gone to the best universities in the country. For example, “Certain Ivy law schools have a track record of producing Supreme Court justices. Four of the eight justices appointed so far in the 21st century earned law degrees from Harvard, and another three graduated from Yale Law School” (Kowarski, 2022). Thus, seven of the eight have law degrees from two of the most selective law schools in the nation. Yet, the Supreme Court of 2022 has six justices with close ties to the ultra-conservative Federalist Society (Feldman, 2021). There is nothing special about this group of Supreme Court justices aside from their conservative ideology.

Intelligence provides a way for us to justify to ourselves and others what we wish to believe, based on the knowledge or supposition of knowledge and beliefs we share with others like ourselves. What makes the current members of the Supreme Court different from past members is not their intellect but their list of associates and friends. A notable example is Harlan Crow, who has provided Supreme Court Justice Clarence Thomas and his wife free luxury vacations worth hundreds of thousands of dollars over a period of many years, including use of his private plane for personal trips (Kaplan et al., 2023). People are convinced, for the most part, not by intellectual firepower, but, rather, by ideological resonance. However, they attribute their ideas to their intellectual skills, regardless of how unimportant those skills may be.

Repetition: Before February 2022, had you ever heard of a “special military operation”? Neither had I; and I have been around a long time! Look at Google, and you now will find 1.21 billion references to this term. Almost all of them are dated after February 24, 2022, when Vladimir Putin announced the so-called “special military operation” against Ukraine.

Russians talk about special military operations a lot because, in true 1984 fashion, Russia has made it against the law to refer to the invasion and the war as either an “invasion” or a “war” (Simon, 2022). So, if one is Russian, one hears over and over again about the “special military operation.” This is part of the firehose of falsehood model described above. One also hears over and over again an imaginary history of Ukraine as an indivisible part of Russia (Greenberg & Jacobson, 2022).

Dictatorships always rewrite history, as China is doing now in Hong Kong (Oung, 2022). The new syllabus in Hong Kong for teaching its history mentions “national security” more than 400 times in 121 pages (Oung, 2022). Dictators such as Xi and would-be dictators such as Trump know that big lies often are easier to sell than small ones (Rosza, 2022).

Multiple but redundant sourcing: For a toxic leader trying to prey on credulous or even not so credulous individuals, it is helpful for a lie to be repeated not only many times but also by multiple sources. Again, this is part of the firehose of falsehood model. One way to ensure that the multiple sourcing comes through is to

make it illegal to say anything other than the government line. In Russia today, for example, spreading “false information” about the military (i.e., true information) is punishable by a 15-year prison sentence (Mathers, 2022). And as we all learned from the case of Alexei Navalny, prison sentences can be indefinitely extended with additional concocted fake charges (Reuters, 2022). Most people do not want to fill the martyr role that Alexei Navalny is filling. So, the independent press has either left the country or shut down or become non-independent. Thus, multiple sources are repeating the same lies. As the January 6 hearings have shown, there have been multiple liars willing to tout Trump’s “Big Lie” about the stolen 2020 US presidential election in the United States. Willing liars are to be found everywhere, and those who are not willing often simply shut up.

Need for affiliation: McClelland (1961/2010) proposed that people have three basic needs: for achievement, power, and affiliation. Toxic leaders are characterized by an overwhelming need for power. Although they may seek to achieve things, these achievements tend overwhelmingly to be ones that will increase their power. When they feel that their power is threatened, they will seek new achievements that will enhance their power. Hence, Putin has overseen the invasion of Ukraine and Xi may very well oversee the invasion of Taiwan.

Their need for power plays into people’s needs for affiliation. For a toxic leader, conformity on the part of followers is paramount. Toxic leaders rule by fear. In Russia or China, if you disobey, you risk jail or death. In the United States, the democratic system has not decayed as much as in Russia or China, so would-be dictators such as Trump use their power to humiliate and suppress their enemies. Trump incites violence, so that one never knows when his followers will come after one. It is not only Trump, unfortunately. In the United States, the extreme left demands conformity as much as the extreme (and not so extreme) right, and people, such as professors who do not conform, risk being humiliated or fired.

People want to belong. They want to be part of something greater than themselves. For some, religion has been a way of finding that something greater. But in some countries, including the US, cynical clerics have subverted and perverted religious doctrine to make it subordinate to a political and ideological agenda and

have had success because the most credulous are so ready to believe (Armaly et al., 2022; Whitehead, 2021). In the case of credulity for conspiracy theories, it might seem odd that credulous people often will go for lies, the more outlandish the better. Why not find things to believe in that everyone knows about? But the advantage of the outlandish is that it makes one special to know about it. The ordinary stuff everyone knows; but the credulous person can feel like part of the deeper cognoscenti by knowing things that other people are, they think, too disconnected or fact-resistant to accept (Jarrett, 2022; Van Prooijen, 2019).

False omniscience: If people keep hearing the same things over and over again from sources they trust, they may come to believe that they have not only reporting, but truth. And this is probably why, in the words of the Fox commentators, the January 6 hearings have not moved the needle. People believe they already know what there is to know. They are foolish in their false sense of omniscience, but view themselves as wise (Sternberg, 2004, 2005).

False omniscience is the opposite of epistemic humility, which is an essential part of wisdom (Grossmann et al., 2020). As one would think people would want to be wise or, at least, appear to be wise, would they want to come across as falsely omniscient? If one looks at leaders around the world, there seem to be more who are falsely omniscient than who are wise. As Lipman-Blumen (1986) and many others (e.g., Örtenblad, 2021; Sternberg, 2021c) have recognized, people are more likely to follow toxic leaders who act falsely omniscient than to follow wise leaders. Sadly, people are attracted precisely to the leaders who often will do the worst by their followers. They are credulous of toxic leaders, ignoring the more balanced and positive views of the wise ones. If things are not going well in one's life, how much easier it is to play the role of the victim and to blame others—the “red meat” of the toxic leader's appeal—than to look for inadequacies in oneself.

How Does a Society Counteract Credulity?

This essay has focused on the truly unfortunate societal conditions that promote credulity and arational thinking. But, as Fiedler (2019) and others (e.g., Stanovich,

2009, 2021) have pointed out, it does not take much for people to think in ways that are less than rational. Misguided thinking is part of everyday life. What can be done to address the credulity that undermines rational and adaptive thinking?

The ways to counteract credulity are scarcely surprising. They are part of any good education.

1. ***Develop and encourage critical thinking:*** Students need to learn to distinguish accurate sources of information from inaccurate ones—to recognize when information is internally consistent and externally correspondent to the truth (Sternberg, 2021b). Critical and rational thinking can help protect people from believing false information (Stanovich, 2009, 2021; Stanovich et al., 2013). IQ in itself will not.
2. ***Develop and encourage wisdom:*** Wise thinking seeks a common good (Sternberg, 2019). It discourages thinking about how one’s own group has a special destiny that leads it, necessarily, to be overlords controlling some other group.
3. ***Provide complete information:*** Censoring information, as Russia, China, and so many other countries do, leads people to think erroneously. In many problem solving studies, people are given complete information needed to solve a problem. In the real world, information too often is deliberately withheld.
4. ***Encourage beneficence:*** There is far too little of it in the world right now. Often, countries are starting to see their nations, and the commonwealth of nations, as a zero-sum game. The United States, for example, is in an advanced state of degeneration, where Republicans and Democrats often seem to place ideological loyalties and loyalties to people like themselves over any kind of national unity. It has become a contest, with “takeovers,” such as the takeover of the Supreme Court of the United States by aggressively ideological conservatives, viewed as a victory by conservatives and as a loss by liberals and many moderates. Only one justice—the Chief Justice, a

conservative—is at all a swing vote, and a generally useless one because of the supermajority of extreme and partisan conservatives. Donald Trump’s brazen attempt to subvert the United States government seems not to have mattered much to many conservatives.

5. ***Remember to teach young people that the truth matters:*** In the United States, at least, and really in a number of other countries, some people have forgotten what it means for something to be true, substituting instead what they wish to be true. It is truth rather than authority that matters in the determination of facts. So, young people must learn to question authority, which is becoming increasingly difficult as more and more of the world succumbs to authoritarian governments and, worse, authoritarian mindsets.

Acknowledgments. I am grateful to Joachim and Marlene for their friendship and hospitality over the years, and for their being so admirable in their caringly looking out for their many friends and colleagues.

References

- Albright, M. (2018). *Fascism: A warning*. Harper.
- Anonymous (n.d.a). *Credulity vs Gullibility – What’s the difference?* WikiDiff.com. <https://wikidiff.com/credulity/gullibility>
- Anonymous (n.d.b). [Search engine result for the query “credulity”]. <https://www.google.com/search?client=firefox-b-l-d&q=credulity>
- Armaly, M. T., Buckley, D. T., & Enders, A. M. (2022). Christian nationalism and political violence: Victimhood, racial identity, conspiracy, and support for the Capitol attacks. *Political Behavior*, 44, 937–960. <https://doi.org/10.1007/s11109-021-09758-y>
- Bick, T., Brüderl, L., Jüttner, A., Dörner, D., Kreuzig, H. W., & Reither, F. (1994). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Huber.
- Bond, C. F., Jr., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, 10(3), 214–234. https://doi.org/10.1207%2Fs15327957pspr1003_2

- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Houghton-Mifflin.
- Chowdhury, M., Hammond, E., Vogt, A., & Wagner, M. (2022, June 23). *Jan. 6 committee holds fifth hearing*. CNN. <https://www.cnn.com/politics/live-news/january-6-hearings-june-23/index.html>
- Codevilla, A. (1993, November/December). *Is Pinochet the model?* Foreign Affairs. <https://www.foreignaffairs.com/articles/chile/1993-12-01/pinochet-model>
- Coppedge, M., Edgell, A. B., Knutsen, C. H., & Lindberg, S. I. (Eds.) (2022). *Why democracies develop and decline*. Cambridge University Press. <https://doi.org/10.1017/9781009086974>
- Desiderio, A. (2021, September 13). *Blinken lays blame on Trump as he defends messy withdrawal from Afghanistan*. Politico. <https://www.politico.com/news/2021/09/13/blinken-lays-blame-on-trump-as-he-defends-messy-withdrawal-from-afghanistan-511677>
- Devega, C. (2021, July 14). *The power of the Big Lie: Why do 30% of Americans cling to Trump's dark fantasy?* Salon. <https://www.salon.com/2021/07/14/the-power-of-the-big-lie-why-do-30-of-americans-cling-to-trumps-dark-fantasy/>
- Diffley, A. (2022, May 25). *Johnson takes responsibility for lockdown-era Number 10 parties but won't quit*. France 24. <https://www.france24.com/en/europe/20220525-partygate-report-blames-leadership-for-illegal-parties-at-boris-johnson-s-office>
- Ecarma, C. (2022, May 6). *Bolsonaro, Brazil's biggest Trump fanboy, apparently setting up his own stolen-election conspiracy*. Vanity Fair. <https://www.vanityfair.com/news/2022/05/jair-bolsonaro-stolen-election-conspiracy>
- Feldman, N. (2021). *Takeover: How a conservative student club captured the Supreme Court*. Pushkin Industries.
- Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical inference. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 1–55). Elsevier Academic Press. <https://doi.org/10.1016/B978-0-12-394293-7.00001-7>
- Fiedler, K. (2019). Metacognitive myopia: Gullibility as a major obstacle in the way of rational behavior. In J. P. Forgas & R. E. Baumeister (Eds.), *The social psychology of gullibility* (pp. 123–139). Routledge. <https://doi.org/10.4324/9780429203787>

- Flynn, J. R. (1987). Massive IQ gains in 14 nations. *Psychological Bulletin*, 101, 171–191. <https://doi.org/10.1037/0033-2909.101.2.171>
- Flynn, J. R. (2012). *Are we getting smarter?* Cambridge University Press. <https://doi.org/10.1017/CBO9781139235679>
- Flynn, J. R. (2016). *Does your family make you smarter? Nature, nurture, and human autonomy*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316576694>
- Forgas J. P., & Baumeister, R. E. (Eds.). (2019). *The social psychology of gullibility*. Routledge. <https://doi.org/10.4324/9780429203787>
- Frensch, P. A., & Funke, J. (Eds.). (1995). *Complex problem solving: The European perspective*. Lawrence Erlbaum Associates.
- Friedland, J. (2023, April 21). *Fox News and its audience became hooked on lies—now they can't break the habit*. The Guardian. <https://www.theguardian.com/commentisfree/2023/apr/21/fox-news-audience-lies-dominion-trial-donald-trump>
- Gardner, F. (2022, February 21). *Ukraine crisis: Five reasons why Putin might not invade*. BBC. <https://www.bbc.com/news/world-europe-60468264>
- Gilbert, D. T. (1991). How mental systems believe. *American Psychologist*, 46(2), 107–119. <https://doi.org/10.1037/0003-066X.46.2.107>
- Goldhagen, D. J. (1997). *Hitler's willing executioners: Ordinary Germans and the Holocaust*. Vintage.
- Graham, D. A. (2022, June 23). *The comment that reveals the depths of the Republican Party's moral collapse*. The Atlantic. https://www.theatlantic.com/ideas/archive/2022/06/rusty-bowers-and-republican-trump-2024-conundrum/661352/?utm_source=newsletter&utm_medium=email&utm_campaign=atlantic-daily-newsletter&utm_content=20220623&utm_term=The%20Atlantic%20Daily
- Greenberg, J., & Jacobson, L. (2022, February 25). *Putin says parts of Ukraine have been Russian since "time immemorial." That's a one-sided view of history*. Poynter. <https://www.poynter.org/fact-checking/2022/was-ukraine-part-of-russia/>
- Grossmann, I., Weststrate, N. M., Ardelt, M., Brienza, J. P., Dong, M., Ferrari, M., Fournier, M. A., Hu, C. S., Nusbaum, H. C., & Vervaeke, J. (2020). The science of wisdom in a polarized world: Knowns and unknowns. *Psychological Inquiry*, 31, 103–133. <https://doi.org/10.1080/1047840X.2020.1750917>

- Jarrett, C. (2022, January 5). *A psychologist explains why people believe in conspiracy theories*. Science Focus. <https://www.sciencefocus.com/the-human-body/conspiracy-theories-belief/>
- Kaplan, J., Elliott, J., & Mierjeski, A. (2023, April 6). *Clarence Thomas and the billionaire*. ProPublica. <https://www.propublica.org/article/clarence-thomas-sotus-undisclosed-luxury-travel-gifts-crow>
- Klass, B. (2017). *The despot's accomplice: How the West is aiding and abetting the decline of democracy*. Oxford University Press.
- Kowarski, I. (2022, April 7). *Where Supreme Court justices earned law degrees*. U.S. News & World Report. <https://www.usnews.com/education/best-graduate-schools/top-law-schools/articles/where-supreme-court-justices-earned-law-degrees>
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. <https://doi.org/10.1037/0022-3514.77.6.1121>
- Levitsky, D., & Ziblatt, S. (2018). *How democracies die*. Crown.
- Mathers, M. (2022, March 4). *Russian дума passes law giving 15-year prison sentences for spreading 'false information' about military*. Independent. <https://www.independent.co.uk/news/world/europe/ukraine-war-latest-russia-law-b2028440.html>
- McClelland, D. C. (1961/2010). *The achieving society*. Martino Fine Books.
- Milbank, D. (2022, June 21). *Those ignoring the Jan. 6 revelations are guaranteeing more violence*. Washington Post. <https://www.washingtonpost.com/opinions/2022/06/21/ignoring-jan-6-hearings-fox-news-violence/>
- MOCKBA. (2022, June 10). *Putin says it is his 'destiny' to 'return and fortify' territories* [Video]. YouTube. Mockba. <https://www.youtube.com/watch?v=3RZU2MKFu8g>
- Mounk, Y. (2018). *The people vs. democracy: Why our freedom is in danger and how we can save it*. Harvard University Press.
- Murray, H. (1938). *Explorations in personality*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195305067.001.0001>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
- Örtenblad, A. R. (Ed.). (2021). *Debating bad leadership: Reasons and remedies*. Springer. <https://doi.org/10.1007/978-3-030-65025-4>

- Oung, A. (2022, June 14). *China rewrites history of Hong Kong with textbooks that deny British rule*. Telegraph. <https://www.telegraph.co.uk/world-news/2022/06/14/china-rewrites-history-hong-kong-textbooks-deny-british-rule/>
- Paul, C., & Mathews, M. (2016). *The Russian “firehose of propaganda” model: Why it works and options to counter it*. Rand. <https://www.rand.org/pubs/perspectives/P E198.html>
- Paulhus, D. L., & Williams, K. M. (2002). The dark triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, 36(6), 556–563. [https://doi.org/10.1016/S0092-6566\(02\)00505-6](https://doi.org/10.1016/S0092-6566(02)00505-6)
- Reuters. (2022, March 22). *Alexei Navalny sentenced to 9 more years in prison after fraud conviction*. The Guardian. <https://www.theguardian.com/world/2022/mar/22/alexei-navalny-13-years-more-jail-fraud>
- Rosza, M. (2022, February 3). *The psychological reason that so many fall for the “Big Lie.”* Salon. <https://www.salon.com/2022/02/03/the-psychological-reason-that-so-many-fall-for-the-big-lie/>
- Serwer, A. (2023, February 19). *Why Fox News lied to its viewers*. The Atlantic. <https://www.theatlantic.com/ideas/archive/2023/02/fox-news-dominion-lawsuit-trump/673132/>
- Shepp, J. (2021, August 15). *The withdrawal from Afghanistan was destined for disaster*. New York Magazine. <https://nymag.com/intelligencer/2021/08/us-withdrawal-from-afghanistan-was-destined-for-disaster.html>
- Simon, S. (2022, March 5). *Russian law bans journalists from calling Ukraine conflict a “war” or an “invasion.”* NPR. <https://www.npr.org/2022/03/05/1084729579/russian-law-bans-journalists-from-calling-ukraine-conflict-a-war-or-an-invasion>
- Stanovich, K. E. (2009). *What intelligence tests miss*. Yale University Press.
- Stanovich, K. E. (2021). *The bias that divides us: The science and politics of myside thinking*. MIT Press.
- Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. *Current Directions in Psychological Science*, 22, 259–264. <https://doi.org/10.1177/0963721413480174>
- Statista. (2022, June 2). *Do you approve of the activities of Vladimir Putin as the president (prime minister) of Russia?* Statista. <https://www.statista.com/statistics/896181/putin-approval-rating-russia/>

- Sternberg, R. J. (2004). Why smart people can be so foolish. *European Psychologist*, 9(3), 145–150. <https://doi.org/10.1027/1016-9040.9.3.145>
- Sternberg, R. J. (2005). Foolishness. In R. J. Sternberg & J. Jordan (Eds.), *Handbook of wisdom: Psychological perspectives* (pp. 331–352). Cambridge University Press. <https://doi.org/10.1017/CBO9780511610486>
- Sternberg, R. J. (2018). Wisdom, foolishness, and toxicity in human development. *Research in Human Development*, 15(3-4), 200–210. <https://doi.org/10.1080/15427609.2018.1491216>
- Sternberg, R. J. (2019). Why people often prefer wise guys to guys who are wise: An augmented balance theory of the production and reception of wisdom. In R. J. Sternberg & J. Glueck (Eds.), *Cambridge handbook of wisdom* (pp. 162–181). Cambridge University Press. <https://doi.org/10.1017/9781108568272.009>
- Sternberg, R. J. (2020a, July 28). *It's time to stem malpractice in STEM admissions*. Inside Higher Ed. <https://www.insidehighered.com/views/2020/07/28/colleges-shouldn-t-use-standardized-admissions-tests-alone-measure-scientific>
- Sternberg, R. J. (2020b). Rethinking what we mean by intelligence. *Kappan*, 102(3), 36–41. <https://kappanonline.org/rethinking-what-we-mean-by-intelligence-sternberg/>
- Sternberg, R. J. (2021a). *Adaptive intelligence: Surviving and thriving in a world of uncertainty*. Cambridge University Press. <https://doi.org/10.1017/9781316650554>
- Sternberg, R. J. (2021b). AWOKE: A theory of representation and process in intelligence as adaptation to the environment. *Personality and Individual Differences*, 182, 111108. <https://doi.org/10.1016/j.paid.2021.111108>
- Sternberg, R. J. (2021c). Why bad leaders? A perspective from WICS. In A. R. Örtengren (Ed.), *Debating bad leadership: Reasons and remedies* (pp. 141–160). Springer. https://doi.org/10.1007/978-3-030-65025-4_8
- Sternberg, R. J. (2022). The vexing problem of dark giftedness. *Gifted Education International*, 39(3), 265–285. <https://doi.org/10.1177/02614294221110459>
- Sternberg, R. J. (2023). Toxic giftedness. *Roeper Review*, 45(1), 61–73. <https://doi.org/10.1080/02783193.2022.2148311>
- Sternberg, R. J., & Frensch, P. A. (Eds.). (1991). *Complex problem solving: Principles and mechanisms*. Lawrence Erlbaum Associates.

- Sternberg, R. J., & Sternberg, K. (2017). Measuring scientific reasoning for graduate admissions in psychology and related disciplines. *Journal of Intelligence*, 5(3), 29. <https://doi.org/10.3390/jintelligence5030029>.
- Sternberg, R. J., Sternberg, K., & Todhunter, R. J. E. (2017). Measuring reasoning about teaching for graduate admissions in psychology and related disciplines. *Journal of Intelligence*, 5(4), 34. <https://doi.org/10.3390/jintelligence5040034>.
- Sternberg, R. J., Todhunter, R. J. E., Litvak, A., & Sternberg, R. J. (2020). The relation of scientific creativity and evaluation of scientific impact to scientific reasoning and general intelligence. *Journal of Intelligence*, 8(2), 17. <https://doi.org/10.3390/jintelligence8020017>
- Sternberg, R. J., Wong, C. H., & Kreisel, A. P. (2021). Understanding and assessing cultural intelligence: Maximum-performance and typical-performance approaches. *Journal of Intelligence* 9(3), 45. <https://doi.org/10.3390/jintelligence9030045>.
- Sternberg, R. J., Wong, C. H., & Sternberg, K. (2019). The relation of tests of scientific reasoning to each other and to tests of fluid intelligence. *Journal of Intelligence*, 7(3), 20. <https://doi.org/10.3390/jintelligence7030020>
- Van Prooijen, J.-W. (2019). Belief in conspiracy theories: Gullibility or rational skepticism. In J. P. Forgas & R. E. Baumeister (Eds.), *The social psychology of gullibility* (pp. 319–332). Routledge.
- Watson, A. (2023, March 28). *Leading cable news networks in the United States in November 2023, by number of primetime viewers*. Statista. <https://www.statista.com/statistics/373814/cable-news-network-viewership-usa/>
- The White House. (2021, July 8). *Remarks by President Biden on the drawdown of U.S. forces in Afghanistan*. <https://www.whitehouse.gov/briefing-room/speeches-remarks/2021/07/08/remarks-by-president-biden-on-the-drawdown-of-u-s-forces-in-afghanistan/>
- Whitehead, A. (2021, May 27). *The growing anti-democratic threat of Christian nationalism in the U.S.* Time. <https://time.com/6052051/anti-democratic-threat-christian-nationalism/>
- Zimmerman, L. (2016). Deception detection. *Monitor on Psychology*, 47(3), 46. <https://www.apa.org/monitor/2016/03/deception>

The Nature of Research on Problem Solving

BENŐ CSAPÓ

University of Szeged, Hungary

Joachim Funke was the chair of the Problem Solving Expert Group in the fifth cycle of the Program for International Student Assessment (PISA). The data-collection phase of this cycle was in the spring of 2012, but outlining the conception and preparing the frameworks and the instruments required more than three years before the actual assessment. The analyses of the data and drafting the report took one more year. The time of the mandate of this group was the period when we had a common task, a few years of intensive work when we ourselves also acted as collaborative problem solvers.

Science is conducted by scientists, individuals who typically work in groups. They are personalities with different characters, backgrounds and skills; people who have specific roles in groups and contribute in many different ways to the success of the collaboration. In most cases, the scientific community and the general public only see the end product of a research and development process. This is usually a publication, in rare cases a license, a new method or a new instrument. The process itself, the overcoming of challenges, the debating of alternatives, the making of crucial decisions, the sharing of work among members, remains usually hidden and only in rare cases do historians of science reconstruct the process and the key events.

If we include the preparations for the first cycle, PISA now has a history of about a quarter of a century. The seven assessment cycles have produced an enormous amount of reports, books, papers and huge databases for secondary analysis. The program has sparked fierce policy debates and has also contributed to the development of education systems in a number of countries. Despite its unprecedented impact, little is known about the work of the expert groups, but the activities of these scholars are interesting, and those responsible for the innovative areas may be particularly adventurous.

In this chapter, I present such a process, focusing on the phase in which Joachim formally played the leading role. However, his scientific work had an impact on the previous period as well, and the group he headed continued and even expanded the joint work further in a period when his activities in PISA had already officially ended. Accordingly, I divide my overview into four parts. First, I outline the context in which the Organisation for Economic Co-operation and Development (OECD) renewed the evaluation of students' knowledge, and from which the assessment of problem solving in a large international assessment program emerged. Then, I describe the period of the first and the second problem solving assessment, finally, the work and the impacts that followed these stages. I am not, of course, attempting to describe the wide-ranging impact of Joachim's work, but to focus on the events and processes—within and beyond PISA—in which I myself was involved. I summarize what I have learned from Joachim, and how I see his role and impact on the research on problem solving.

The Role of OECD in Renewing the Assessment of Students' Knowledge

The OECD aims to support social and economic development in all countries of the world, as summarized in a current statement: “Our goal is to shape policies that foster prosperity, equality, opportunity and well-being for all.”¹ The

¹ See <https://www.oecd.org/about/>, retrieved December 20, 2022.

OECD has a large statistical database which allows sophisticated analyses and modeling the forces that drive social progress and economic growth. As became clear from economic research, education is a main engine of development and the OECD therefore collected data about schooling as well. However, the data that were available earlier about the education in the member countries did not really characterize the quality of education.

It is well known that there are countries in which students acquire more knowledge of a higher quality in a shorter period of time than students in other countries with a weaker school system. Thus, simple school enrolment data do not say much about how a school system (or more broadly, a society as a whole) is able to create new knowledge, the type of knowledge that really supports development, prosperity, and well-being. This demand triggered the construction of a measurement system that can provide a sufficiently detailed description of the performance of school systems. PISA was designed for this purpose, assessing knowledge with the most advanced instruments and completing data-collection with a broad range of background variables to offer a sound basis for policy-relevant analyses.

An essential prerequisite for creating such an evaluation system is to clarify what kind of knowledge young citizens of developed societies need for preparing a successful private and professional life. This question was explored by several OECD expert bodies. These efforts included the *Definition and Selection of Key Competencies* (DeSeCo) project. The initial ideas were discussed in the framework of a series of workshops and two symposia in 1999 and 2002. The conclusion of this work was, in brief, that in order to live a successful life, citizens of a well-functioning society have to be able to work in heterogeneous groups, act autonomously, and use tools such as language or technology interactively. The results of DeSeCo were published in two volumes (Rychen & Salganik, 2001, 2003) and influenced several further programs.

It was also obvious that PISA had to follow the common standards of international assessments and so had to include the three main domains that are usually assessed in large-scale national and international programs: reading, mathematics, and science. However, the conception of knowledge was renewed in these areas as

well, by extending the concept of *literacy*. Reading literacy, mathematical literacy and scientific literacy were defined so that they included how well students understood key concepts, how they mastered certain processes, and how they could apply their knowledge in real-world situations. In contrast to some previous programs, PISA did not aim at examining how successfully students mastered the school curriculum, but rather how young people were able to apply their knowledge and skills in order to successfully cope with life's challenges. These principles were set out before the first assessment (OECD, 1999), and are slightly revised before each survey cycle in the assessment frameworks, which are published well before the actual data collection.

The goal of PISA is to monitor the achievements of the educational systems by assessing various aspects of knowledge of 15-year-old students. The topics of the measurements are aimed at being aligned with social changes and reflecting new challenges. The realization of this goal does not always fit into the three traditional domains; there are several interesting and relevant novel fields which cannot easily be included into their assessments. Furthermore, a basic requirement is that the trends of change in each country should be clearly seen from the results. This expectation can be met only if the measuring instruments remain unchanged. Thus, it is not possible to significantly change the tests from cycle to cycle: the instruments used for the three main domains must remain stable.

As the assessments of reading, mathematics, and science were used in this way to outline trends, further domains had to be added to assess novel aspects of students' competencies to reflect changing expectations. Therefore, a fourth, so called *innovative* domain is assessed in every PISA cycle. This innovative domain can always be a new one without any restriction concerning its content and the method of measurement.

By adding such an innovative domain, PISA faced a new challenge: agreeing on an original topic which is interesting, relevant for policy making, and measurable in international contexts. A further expectation is that the innovative domain should measure something, which cannot be predicted from the data of the other domains. In other words, the innovative domain should add a new measurement dimension,

so that it can distinguish countries which are similar according to the achievements in reading, mathematics, and science.

In the first assessment in 2000, the innovative domain explored students' approaches to learning by examining their learning strategies, their motivation and their confidence in their own learning abilities. A questionnaire using a set of scales was used to generate a picture about the way students learn in the participating countries and the results revealed several so far unknown aspects of education. For example, they have shown that there are large differences between countries in the frequency with which students use rote memorization versus elaborative learning strategies, and they help explain some of the differences in cognitive test performances (Artelt et al., 2003).

The First PISA Problem Solving Assessment—The Early Period of Cooperation

The origin of assessing problem solving in PISA can be traced back until 1999, when the Network A of the OECD (a group responsible for the development of indicators of learning outcomes) commissioned a paper on the possibility of assessing problem solving within PISA. A group of experts were invited for further elaborating this idea under the umbrella of Network A. These experts (Network A Problem Solving Group) met for the first time in Chicago in April 2000, around the same time when the data collection of the first cycle took place. This date also indicates that the preparation started well before the actual survey and it took three years. This group consisted of five members: John Dossey, chair, Illinois State University (USA); Benő Csapó, University of Szeged (Hungary); Ton de Jong, University of Twente (The Netherlands); Eckhard Klieme, German Institute for International Educational Research (Germany) and Stella Vosniadou, National and Capodistrian University of Athens (Greece). Two further working meetings were organized within Network A, also attended by Eugene Owen, a representative of OECD. The initial results were presented to the Board of Participating Countries (BPS) in October 2000 at its meeting in Bremen. After some discussion the con-

ceptual framework was accepted by the BPS, and a decision was made to include the assessment of cross-curricular problem solving within PISA 2003.

Subsequently, the Problem Solving Expert Group (PEG) was officially formed and its first formal meeting took place in Brussels in February 2001. The research experience of the members of PEG covered a broad range of fields in cognition, cognitive development, and educational testing. Although our fields of research were related to some aspects of problem solving, none of us was a specialist in research on problem solving. When I asked the colleagues working in the OECD, why this group was put together in this way, they explained that by this composition the probability of non-traditional, creative solutions could be increased. Furthermore, designing and implementing large-scale assessment in international context, where the participating countries are culturally diverse, require other approaches than usual in psychological research. Later on, we realized the challenges that emerged from this expectation.

Problem solving is one of the broadest intellectual processes known, and it is also one of the most intensively researched cognitive constructs. It is used almost everywhere, and has many different definitions and interpretations. So the first task of the expert group was to narrow the conception of problem solving and define it in a measurable format to fit the general aims of the PISA assessments. Looking back to that period, I see three main challenges.

First, we had to study the relevant literature on problem solving in depth. Besides reading or re-reading several classical works on problem solving, this was the phase when I realized the importance of Joachim's contribution to the field. When the group had decided, that the complex problem solving (CPS) paradigm will be the main direction of the approach, several aspects and characteristics of CPS were adopted from his publications, for example, the concept of complexity and the difference between analytical and complex problem solving (Funke, 1991, 1995; Frensch & Funke, 1995). As I was interested in computer-based assessment, I was especially impressed with his works on the use of computers in the assessment of CPS (Funke, 1998). For a while, the expert group considered computer-based assessment as well, but as the conditions were not present in many countries, this

idea was dropped in earlier assessment cycles. (For the first conceptualization see Dossey et al., 2000).

A second challenge arose from the fact that each main domain (reading, mathematics and science) already had problem solving components as part of their literacy definition in the framework (see OECD, 1999). The potential overlap was especially large for mathematical literacy. So the expert group had to pay attention to avoid the possible overlap with the other domains and clearly separate CPS from them. Therefore, the content of the assessment was moved from familiar to unfamiliar settings, the context from school-based topics to real-world, and the complexity increased from single discipline to multiple applications. For harmonizing the assessment of CPS with the other domains, the chairs of the other PISA expert groups, Jan De Lange (Utrecht University, The Netherlands, Mathematics Expert Group), Wynne Harlen (University of Bristol, United Kingdom, Science Expert Group) and Irwin Kirsch (Educational Training Service, United States, Reading Expert Group) joined our core PEG.

Third, we had to communicate our outcomes to others. First of all, we had to report about our work to the PISA BPC. This task was usually done by John Dossey and Eckhard Klieme, as they were most familiar with problem solving research. Furthermore, the assessment of CPS needed to be explained to a broader scientific community, and finally we had to be prepared to clarify the results to the stakeholders, policy makers and to a lay audience. As this was the first time that a cross-curricular competency was measured in an international assessment program, it was not an easy job. In the first phase of our communication efforts, we realized that evaluating problem solving is by no means a self-evident and understandable idea. Although developing problem solving is a declared goal of schooling, there is no subject in the curriculum called problem solving. A question that has often been asked since the beginning of PISA is: Is it right to assess something that is not taught in school? Representatives of some countries feared (sometimes rightly) that such a survey would highlight the weaknesses of their school systems. And we have had to argue again and again that PISA (and in this particular case the CPS assessment) is not a competition for better rankings, but a tool for identifying

needs and opportunities for improvement. Beyond arguing for the relevance of CPS assessment, we also had to deal with the validity and reliability of the prepared instruments, which also raised unexpected challenges.

The expert group, in agreement with all stakeholders, created a balanced definition that expressed all major aspects of CPS relevant to PISA assessment:

Problem solving is an individual's capacity to use cognitive processes to confront and resolve real, cross-disciplinary situations where the solution path is not immediately obvious and where the literacy domains or curricula areas that might be applicable are not within a single domain of mathematics, science or reading. (OECD, 2003, p. 156).

The phases of problem solving described in the framework were in line with the well-known steps that Polya (1945) identified decades ago (understanding, characterizing, and solving the problem, then reflecting on the solution and communicating it), and this familiarity also helped to improve the acceptance of the CPS assessment. The results clearly indicated that CPS was different enough from the other domains and indicated interesting differences between the participating countries which were new compared to the results of the other domains (OECD, 2004). This was a promising direction for the future innovative domain.

The Story of the PISA 2012 Dynamic Problem Solving Assessment

When the preparation processes for the PISA 2012 assessment cycle started, I was a member of the PISA Governing Boards (PGB), more exactly, I was the vice chair of it. Although there was not a strict formal share of responsibilities within the executive group (the chair and the three vice chairs), due to my research background, I was asked to elaborate a proposal for the innovative domain of the 2012 assessment cycle. As in 2003, when mathematics was the main assessment domain, and complex problem solving was measured in parallel with it, it was an obvious idea that some kind of problem solving should be measured again.

Nevertheless, it was not required to connect it to the previous CPS assessment, therefore I was encouraged to incorporate all current developments of the field into the proposal.

The document I prepared (Csapó, 2008) was discussed and approved by the PGB and became an Annex of the Terms of Reference for the 2012 assessment (see Annex E of OECD, 2008). In this proposal I suggested a new definition of problem solving, and, for maximizing its added value, I proposed that it should be “further distanced and better distinguished from the three core literacy domains” (OECD, 2008, p. 67). Computer delivery was also a crucial component of the proposal. This idea was not new in PISA at that time, as in 2006 the Computer-based Assessment of Scientific Literacy (CBAS) was already an optional part. Nevertheless, only three countries (Denmark, Iceland and Korea) participated in that optional assessment, and the final report about it appeared only years later (OECD, 2010). Furthermore, the second computer-based assessment in PISA, the online reading assessment in 2009, was still in preparation (see OECD, 2011), so there was little evidence that computer delivery would be possible in all participating countries. On the other hand, the 2012 PISA was planned as a breakthrough in terms of computer delivery: a computerized version of the main domains (reading, mathematics, and science) was also offered for all countries.

The problem solving expert group began its work with this background. The situation was somewhat controversial, as on one hand the expert group was expected to design an original assessment that would go beyond both the previous CPS assessment and the already planned computer delivery of the main assessment domains. On the other hand it seemed unlikely that such an assessment would be possible on an international scale, taking into account the diversity of the participating countries.

The first meeting of the Problem Solving Working Party (as it was called in the meeting documents) for outlining the new problem solving assessment was held in Santa Barbara (Richard Mayer’s home university) in July, 2009. This meeting was very well prepared, and the materials distributed to the participants summarized the goals of PISA, and described in details the former problem solving assessment;

including the background processes and of what we have learned from the previous works. The main purpose of this meeting, chaired by Ray Adams, was to establish a common understanding of the conceptual framework of the problem solving assessment. The presentations, an overview of PISA in general (by Ray Adams), the PISA 2003 problem solving (John Dossey), the background of the PISA 2012 problem solving (Ray Adams) and the technical possibilities of the computer-based delivery (Maurice Walker) were followed by intensive discussion.

However, there were a large number of dilemmas and the participants were divided on several issues. For example, whether the 2012 problem solving assessment should be connected to the 2003 CPS assessment, and whether it should be linked to the Programme for the International Assessment of Adult Competencies (PIAAC), which had similar tasks. A further question remained, which had already been raised in the 2003 assessment, as to how problem solving could be a construct distinct enough from the main literacy domains if they were also assessed computer-based and included (domain-specific) problem solving components.

To come to an agreement, another meeting was organized shortly after the first one in September 2009, also in Santa Barbara. To broaden the scope of the possibilities, Kathleen Scalise and Marilyn Binkley presented an overview of the Assessment and Teaching of 21st Century Skills project (for the results of the first phase of this initiative, see Griffin et al., 2012). As the initial proposals seemed too traditional, I also gave a presentation, arguing that we should further decrease the overlap between problem solving and the other three domains and increase the novelty and added value of this assessment. I argued that it should be as innovative as possible, as a theoretical construct as well as in the operationalization of its assessment. It seemed to me that connecting it to the 2003 CPS assessment would constrain the possibilities; therefore I also proposed to give up the idea of such a connection and to construct an entirely different assessment. I argued that success in the assessment of problem solving would determine future directions and developments in this area. I also mentioned that dynamic problems seemed to be a fruitful area to pursue. In the discussion that followed my presentation,

it became clear that the extent of innovation should constantly, but gradually be increased across the assessment cycles.

By that time, I was familiar with the research of Joachim Funke in dynamic problem solving (e.g., Funke, 1988, 1992, 1993, 2001). I was especially interested in the relationship between the knowledge acquisition phase of dynamic problem solving and inductive reasoning. I also knew him personally; I met him and his research group several times at the annual workshops (colloquia) of the German DFG Priority Program “Competence Models for Assessing Individual Learning Outcomes and Evaluating Educational Processes”. I was invited to be a “critical friend” of the program, and my task was to read and critically comment on the materials of the participating research groups.

So, in that second meeting in Santa Barbara, I argued in favor of Funke’s approach, in part because the reasoning processes involved are more visible than in traditional problem-solving tasks. However, I was not sure whether we could go this far in a large-scale assessment and whether dynamic tasks could be implemented on computers so that such an assessment could be carried out in all participating countries. The arguments of colleagues with experience in international assessment seemed convincing, and I agreed that we needed to keep the risks at an acceptable level. Luckily, Detlev Leutner also contributed to the discussion and had a number of encouraging remarks when he explained that various aspects of problem solving had already been assessed in Germany in several contexts. He argued convincingly that dynamic problem solving was the best option (see also Klieme et al., 2005; Leutner & Wirth, 2005). Perhaps his thoughtful, professional argument was the final moment that decided the debate, and the majority of those present accepted that dynamic problem solving should be the innovative domain. At the same meeting, in recognition of his prominent role in problem solving research and his pioneering role in dynamic problem solving, it was decided that Joachim should be invited to chair the expert group.

From that point onwards, the work followed the protocol established by PISA. An expert group was formally appointed, chaired by Joachim Funke (Heidelberg University) with seven core members: John Dossey (Illinois State University,

United States), Arthur Graesser (University of Memphis, United States), Detlev Leutner (Duisburg-Essen University, Germany), Romain Martin (Université de Luxembourg FLSHASE, Luxembourg), Richard Mayer (University of California, United States), Ming Ming Tan (Ministry of Education, Singapore) and Benő Csapó (University of Szeged, Hungary). The first task of the group was to elaborate the assessment framework, creating a working definition, describing the organization of the domain, and specifying the characteristics of the assessment. After carefully discussing all views, a balanced definition was conceived based on Joachim's former work (e.g., Funke, 2010), taking into account the needs and limitations of a large-scale assessment:

Problem-solving competency is an individual's capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious. It includes the willingness to engage with such situations in order to achieve one's potential as a constructive and reflective citizen. (OECD, 2013, p. 122).

Of course, this simple, quite general, and easy to accept definition did not mirror the challenges the expert group faced mentioned in the previous sections. It still remained open to what extent the assessment could be based on the conception of dynamic problem solving, more specifically, on Joachim's approach. Some members proposed developing a "purely dynamic" but somewhat risky instrument, while the broader group of stakeholders preferred to remain on the safer side and following the most traditional and already tried-out tests. The difficulty of including dynamic problem solving was resolved by introducing the aspect of the "nature of the problem situation" and distinguishing static and interactive problem situations. By explaining the interactive situations, the crucial element, the first phase of the dynamic problem solving was described: "some exploration or experimentation must be done to acquire the knowledge necessary to control the device" (OECD, 2013, p. 125). Nevertheless, static problem situations were also described in the framework, leaving space for the more traditional approaches.

The expert group held a number of meetings, many of them with exciting debates, and Joachim guided the group through the risks and challenges with his calm professionalism. The final output of our work, the intellectual basis of the actual assessment, was published in the PISA results volume (OECD, 2014). The dynamic problems were described as “[i]nteractive: not all information is disclosed; some information has to be uncovered by exploring the problem situation” while there remained items that were “[s]tatic: all relevant information for solving the problem is disclosed at the outset” (OECD, 2014, p. 31). In the final instrument, there were five units with static items and altogether ten interactive units. These interactive items embodied ideas in a large-scale assessment which Joachim suggested decades ago (e.g., Funke, 1991, 1993, 2001). Six of them were based on the finite-state automata (Buchner & Funke, 1993; Funke, 2001) and four of them on linear structural equations (Funke, 1991, 1992; Greiff & Funke, 2009).

The actual assessment phase in 2012 went smoothly; the computerized tests were administered altogether to 85,000 students from 44 countries and economies. Neither the nature of dynamic problem solving, nor the computer delivery caused any specific difficulty, so the initial concerns proved unfounded (or, the foreseen risks were successfully managed). The 2012 problem solving assessment was, by any measure, extremely successful. It revealed dimensions of differences between countries that had not yet surfaced in previous surveys. It also paved the way for the 2015 collaborative problem-solving assessment, which also used interaction with a computer-simulated system and included the dimension of group dynamics in the measurement.

The Post-PISA-2012 Era—Long-Term Impacts

The meetings of the expert group were always an exciting experience. Problem solving, which never seemed like work, continued with intensive correspondence between official meetings. As with the personal meetings, problem solving was not limited to the time on the agenda, but began at breakfast and continued over dinner until late in the evening. The meetings were held in such memorable locations as

Boston, Melbourne, Budapest and Heidelberg, but even these locations could not distract the group from generating new and fresh ideas.

The main challenge in this process was to find a balance between the ideas of the participants and the possibilities of the international survey. The creative atmosphere maintained by Joachim inspired such a mass of ideas that they could no longer be squeezed into official documents and new opportunities had to be created.

One such opportunity was that we participated in many conferences where we could present the results of joint activities or mutually inspired research without formal constraints. This cooperation, which was separate from the PISA activities, also went beyond the formal commitments and continued even after the publication of the volume summarizing the PISA results. Such a meeting took place, among others, at the AERA conference in New Orleans, at the EARLI conference in Exeter, and the Szeged Workshop on Educational Assessment (SWEE), held in Szeged every spring between 2009 and 2016. On one occasion, in 2011, Joachim was co-chair of SWEE, which means that he played a crucial role in the compilation and execution of the program. The most memorable event probably was the “Celebrating Problem Solving” conference in November 2015, which was organized on the occasion of the University of Szeged awarding Joachim Funke the title of Doctor Honoris Causa.

Several people in this group then invited their colleagues to these meetings, until eventually a significant informal network of researchers interested in problem solving was formed. Finally, we decided to publish a volume of the work of this intellectual community. This turned out to be a unique phenomenon among the PISA expert groups, in that researchers in a field were kept together by a common interest even years after the end of their mandate. Since the inspiration for the book came from the joint activities carried out under the auspices of the OECD, the organization undertook to publish the volume, which I edited together with Joachim (Csapó & Funke, 2017).

The collaboration with Joachim resulted in numerous other long-lasting impacts on the professional career and research orientation of the participants. To mention

only those in Szeged, we developed an online assessment system, the eDia (Csapó & Molnár, 2019). The main function of eDia is to provide diagnostic information for primary school students through regular assessment in reading, mathematics, and science. It has two main parts: One is an online platform that supports all aspects of testing, from writing sophisticated items, compiling and administering tests, analyzing the results, and providing detailed visualized feedback to students and their teachers. The other part is an item bank containing tens of thousands of items.

The platform can be used for any other assessment and we have implemented the assessment of dynamic problem solving in this system as well. After the abstract simulation works, concrete items are easy to create. The background story of the problems and other parameters (the number of variables involved and the relationships between them) can be freely changed. So we can run assessments with a large number of items (which Joachim predicted in the early 1990s) and we can do so in multiple contexts. For example, we measure our students' readiness for college at the beginning of their freshman year. We have tested the predictive power of several cognitive constructs, including dynamic problem solving (Csapó & Molnár, 2017). By assessing problem solving with this system, we can separate the results of the two phases of dynamic problem solving, knowledge acquisition and knowledge application. Finally, we were also able to provide empirical evidence that knowledge acquisition is more strongly correlated with other cognitive variables than knowledge application.

Some Concluding Personal Remarks

After reviewing Joachim's scientific work and our joint activities, I realized how many similarities there are in our professional careers, despite the many differences in the circumstances of our personal lives. We were born in the same year, and although we took very different paths, we became full professors in the same year. We spent the first 35 years of our lives in different political systems on two different sides of the Iron Curtain. Nevertheless, we seem to have developed a

similar value system and we think the same way about many things in life and science. The language in which we exchange our thoughts is not our mother tongue, but that has never prevented us from understanding each other perfectly.

In 1989, when I was a Humboldt Scholar in Bremen, we witnessed the fall of the Berlin Wall from the same side. It was the year that brought about a change of historic significance in both our countries. Since then, the changed world order has allowed us to be involved in the same scientific processes and to participate in the same professional and scientific events in the second 35 years of our lives.

To emphasize the importance of educational measurement, we often quote Lord Kelvin, who said “if you cannot measure it, you cannot improve it”. The number of publications is often cited as an indicator of a researcher’s success. Citations can be counted; impact can thus be measured by the number of citations or by other indicators derived from the number of citations. Looking at Joachim’s scientometric numbers, we can say that he is a successful researcher with a high impact. However, the old English saying (sometimes attributed to Einstein) that “the things that really count cannot be counted” shows the other side of the coin. This is the case for a difficult to define concept: the respect and recognition a scientist receives from colleagues, fellow researchers, and students. Joachim has earned this recognition.

Looking back on our collaboration over the past decades, I understand not only the nature of problem-solving research, but also the nature of academic friendship; or perhaps the nature of friendship in general, the nature of friendship without adjectives. With this chapter, I add my voice to the immeasurable respect that Joachim is held in by his colleagues and friends.

References

- Artelt, C., Baumert, J., Julius-McElvany, N., & Peschar, J. (2003). *Learners for life. Student approaches to learning. Results from PISA 2000*. OECD Publishing. <https://doi.org/10.1787/9789264103917-en>

- Buchner, A., & Funke, J. (1993). Finite-state automata: Dynamic task environments in problem-solving research. *The Quarterly Journal of Experimental Psychology*, 46(1), 83–118.
- Csapó, B. (2008). *Proposal for assessing general problem solving competence in PISA 2012*. Unpublished.
- Csapó, B., & Funke, J. (Eds.). (2017). *The nature of problem solving. Using research to inspire 21st century learning*. OECD Publishing. <https://doi.org/10.1787/9789264273955-en>
- Csapó, B., & Molnár, G. (2017). Potential for assessing dynamic problem-solving at the beginning of higher education studies. *Frontiers in Psychology*, 8, 2022. <https://doi.org/10.3389/fpsyg.2017.02022>
- Csapó, B., & Molnár, G. (2019). Online diagnostic assessment in support of personalized teaching and learning: The eDia system. *Frontiers in Psychology*, 10, 1522. <https://doi.org/10.3389/fpsyg.2019.01522>
- Dossey, J., Csapó, B., de Jong, T., Klieme, E., & Vosniadou, S. (2000). Cross-curricular competencies in PISA. Towards a framework for assessing problem-solving skills. In INES (Eds.), *The INES Compendium. Contributions from the INES Networks and Working Groups*, (pp. 19–41). OECD Publishing.
- Frensch, P. A. & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective*. (pp. 3–25). Lawrence Erlbaum.
- Funke, J. (1988). Using simulation to study complex problem solving: A review of studies in the FRG. *Simulation & Games*, 19(3), 277–303. <https://doi.org/10.1177/0037550088193003>
- Funke, J. (1991). Solving complex problems: Exploration and control of complex systems. In R. J. Sternberg & P. A. Frensch (Eds.), *Complex problem solving: Principles and mechanisms* (pp. 185–222). Lawrence Erlbaum Associates.
- Funke, J. (1992). Dealing with dynamic systems: Research strategy, diagnostic approach and experimental results. *German Journal of Psychology*, 16(1), 24–43.
- Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. In G. Strube & K.-F. Wender (Eds.), *The cognitive psychology of knowledge* (pp. 313–330). Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)62663-1](https://doi.org/10.1016/S0166-4115(08)62663-1)

- Funke, J. (1995). Experimental research on complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving: The European perspective*. (pp. 243–268). Lawrence Erlbaum.
- Funke, J. (1998). Computer-based testing and training with scenarios from complex problem-solving research: Advantages and disadvantages. *International Journal of Selection and Assessment*, 6(2), 90–96. <https://doi.org/10.1111/1468-2389.00077>
- Funke, J. (2001). Dynamic systems as tools for analyzing human judgment. *Thinking and Reasoning*, 7(1), 69–89. <https://doi.org/10.1080/13546780042000046>
- Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, 11(2), 133–142. <https://doi.org/10.1007/s10339-009-0345-0>
- Greiff, S., & Funke, J. (2009). Measuring complex problem solving: The MicroDYN approach. In F. Scheuermann & J. Björnsson (Eds.). *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (pp. 157–163). Office for Official Publications of the European Communities. <https://10.11588/heidok.00015502>
- Griffin, P., McGaw, B., & Care, E. (Eds.). (2012). *Assessment and teaching of 21st century skills*. Springer. <https://doi.org/10.1007/978-94-007-2324-5>
- Klieme, E., Leutner, D., & Wirth, J. (Eds.) (2005). *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA 2000 Studie*. VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-322-85144-4>
- Leutner, D., & Wirth, J. (2005). What we have learned from PISA so far: A German educational psychology point of view. *KEDI Journal of Educational Policy*, 2(2), 39–56.
- OECD. (1999). *Measuring student knowledge and skills: A new framework for assessment*. OECD Publishing. <https://doi.org/10.1787/9789264173125-en>
- OECD. (2003). *The PISA 2003 assessment framework. Mathematics, reading, science and problem solving knowledge and skills*. OECD Publishing. <https://doi.org/10.1787/9789264101739-en>
- OECD. (2004). *Problem solving for tomorrow's world. First measures of cross-curricular competencies from PISA 2003*. OECD Publishing. <https://doi.org/10.1787/9789264006430-en>

- OECD. (2008). *Terms of reference for the PISA 2012 assessment: Draft framework. 26th meeting of the PISA Governing Board, The Hague, 2–5. November 2008. EDU/PISA/ GB(2008)26/REV1*. [https://one.oecd.org/document/EDU/PISA/GB\(2008\)26/REV1/en/pdf](https://one.oecd.org/document/EDU/PISA/GB(2008)26/REV1/en/pdf)
- OECD. (2010). *PISA Computer-based assessment of student skills in science*. OECD Publishing. <https://doi.org/10.1787/9789264082038-en>
- OECD. (2011). *PISA 2009 results: Vol. 6. Students on line. Digital technologies and performance*. OECD Publishing. <https://doi.org/10.1787/9789264112995-en>
- OECD. (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. OECD Publishing. <https://doi.org/10.1787/9789264190511-en>
- OECD. (2014). *PISA 2012 results: Vol. 5. Creative problem solving: Students' skills in tackling real-life problems*. OECD Publishing. <https://doi.org/10.1787/9789264208070-en>
- Polya, G. (1945). *How to solve it*. Princeton University Press. <https://doi.org/10.1515/9781400828678>
- Rychen, D. S., & Salganik, L. H. (Eds.). (2001). *Defining and selecting key competencies*. Hogrefe & Huber.
- Rychen, D. S., & Salganik, L. H. (Eds.). (2003). *Key competencies for a successful life and a well-functioning society*. Hogrefe & Huber.
- Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving – More than reasoning? *Intelligence*, 40(1), 1–14. <https://doi.org/10.1016/j.intell.2011.11.003>

Die vier Wellen der psychometrischen Forschung zu komplexen Problemen

MATTHIAS STADLER

Ludwig-Maximilians-Universität München

Als ich während meines Studiums zum ersten Mal mit den Arbeiten von Joachim Funke in Berührung kam, war ich sofort begeistert. Seine Ansätze zur Lösung komplexer Probleme, insbesondere seine Arbeiten zur Schneiderwerkstatt, haben mich fasziniert wie kaum ein anderer Inhalt meines Studiums. Seitdem verfolge ich seine Arbeiten aufmerksam und es ist mir eine große Ehre, zu dieser Festschrift anlässlich seines 70. Geburtstags beizutragen. In diesem kurzen Text möchte ich die Entwicklung der deutschen Forschung zur Lösung komplexer Probleme aus meiner Sicht erläutern. Dabei habe ich diese Entwicklung in vier Wellen unterteilt, die jeweils einen grundlegenden Wandel im Verständnis komplexer Probleme und in der methodischen Herangehensweise an die Erforschung menschlichen Problemlösens darstellen.

Ich beginne mit den Arbeiten von Dietrich Dörner, auf dessen bahnbrechenden Studien unser Forschungsgebiet beruht. Als nächstes werde ich die Kritik an dieser Forschung zusammenfassen und zeigen, wie Joachim Funke versucht hat, dieser Kritik zu begegnen. Danach werde ich zeigen, wie Samuel Greiff die Forschung von Joachim Funke auf die große Bühne der internationalen Large-Scale-Studien geführt hat, und schließlich werde ich meine aktuelle Sichtweise auf komplexes Problemlösen erläutern.

Da dies kein umfassender Übersichtsartikel sein soll, werde ich keine dieser Perspektiven in ausreichender Form darstellen können. Darüber hinaus werde ich mich fast ausschließlich auf den „psychometrischen Ansatz“ konzentrieren, der jedoch nicht die einzige Betrachtungsweise des komplexen Problemlösens ist (Dörner & Funke, 2017). Nichtsdestotrotz hoffe ich, dass dieser kurze Überblick den einen oder die andere zum Nachdenken über die eigene Position in der Forschung zum komplexen Problemlösen anregt oder als Grundlage für weiterführende Diskussionen dienen kann.

Die Operative Intelligenz

Was ist Intelligenz? Laut dem Board of Scientific Affairs der American Psychological Association ist Intelligenz „die Fähigkeit, komplexe Sachverhalte zu verstehen, sich effektiv an die Umwelt anzupassen, aus Erfahrungen zu lernen, verschiedene Formen des Denkens zu nutzen und Hindernisse durch Nachdenken zu überwinden“ (Neisser et al., 1996). Aber messen herkömmliche Intelligenztests diese Fähigkeit? Muss man sich wirklich an die Umwelt anpassen, um eine Matrizenaufgabe zu lösen, wie sie häufig in klassischen Intelligenztests vorkommt, oder muss man aus Erfahrungen lernen, um Zahlenreihen zu vervollständigen? Eher nicht.

Diese Diskrepanz zwischen theoretischer Definition und empirischer Operationalisierung des Konstrukts Intelligenz wurde von Dietrich Dörner in den 1970er- und 1980er-Jahren kritisiert (z. B. Dörner, 1986). Ausgehend von früheren Befunden zum menschlichen Problemlösen in mehrdeutigen Situationen (z. B. Toda, 1962) argumentierte er, dass echte (ökologisch valide) Probleme im Gegensatz zu den in Intelligenztests zu lösenden Problemen dynamisch sind, d. h. sich über die Zeit oder durch Interventionen verändern (Dynamik). Darüber hinaus sind die zu erreichenden Ziele oft vielschichtig (Polytelie) und der Weg dorthin nicht immer offensichtlich (Intransparenz), da zwischen den Elementen des Problems Abhängigkeiten bestehen (Interdependenz). Um die Fähigkeit, solche komplexen Probleme zu lösen, beobachtbar zu machen, entwickelte Dörner hochkomplexe Simulationen, in denen die Proband*innen Städte oder sogar ganze Länder zu

verwalten hatten. Er fand heraus, dass sogar Expert*innen Schwierigkeiten hatten, diese Aufgaben zu bewältigen, es aber durchaus Eigenschaften gab, die gute und schlechte Problemlöser*innen voneinander unterschieden. Brisanterweise war dabei nicht die konventionell gemessene Intelligenz von großer Bedeutung, sondern die emotionale Reaktion. Der Unterschied liegt also in den unterschiedlichen emotionalen Reaktionen auf die Schwierigkeiten und Anforderungen des Problems:

Während gute Problemlöser in schwierigen Situationen Ruhe bewahren und versuchen, die Ursachen der Schwierigkeiten bzw. ihrer eigenen Unfähigkeit, mit ihnen umzugehen, zu ergründen, geraten schlechte Problemlöser in Versagensängste und damit in Stress, der zu einem zerfahrenen, desorganisierten Verhalten führt. Die völlig Erfolglosen reden viel, tun aber wenig Sinnvolles. (Dörner et al., 1983)

Dieser Generalangriff auf die etablierte Intelligenzforschung konnte natürlich nicht unwidersprochen bleiben. Die Forscher*innen um Heinz-Martin Süß brachten gleich eine ganze Reihe von Kritikpunkten an Dörners Forschung vor. Es gäbe keine optimale Lösung und potenziell ausweglose Situationen. Außerdem seien die Probleme viel zu schwierig und aufgrund ihrer Länge nicht mit mehreren Aufgaben zu testen. Vor allem aber zeigten sich in einer sehr groß angelegten empirischen Studie nur geringe Zusammenhänge zwischen den Leistungen bei verschiedenen komplexen Problemen, die fast vollständig verschwanden, wenn die Verarbeitungskapazität und das domänenspezifische Vorwissen kontrolliert wurden (Süß, 1999). Die Schlussfolgerung aus dieser theoretischen und empirischen Kritik war, dass es kein eigenständiges Konstrukt *komplexes Problemlösen* gibt (Wilhelm, 2009).

Fokus auf psychometrische Qualität

Damit kommen wir schließlich zu Joachim Funke und seiner elementaren Rolle bei der Neuinterpretation der Forschung zum komplexen Problemlösen. Selbst

kein Doktorand von ihm, aber laut seinem sehr unterhaltsamen Blog ein Fan von Dietrich Dörner, war Joachim Funke schnell von der Idee des komplexen Problemlösens begeistert. Auf die heftige Kritik an Dörners frühen Arbeiten reagierte er mit zwei wesentlichen Veränderungen des Ansatzes zur Messung komplexer Problemlösefähigkeit. Zunächst entwickelte er Möglichkeiten, komplexe Probleme zu formalisieren. Während Dörners Simulationen noch sehr idiosynkratisch waren, um reale Probleme möglichst genau nachzubilden, schlug Funke vor, Probleme als lineare Strukturgleichungen oder finite Automaten zu beschreiben (Funke, 2001). Diese wesentlich einfachere mathematische Beschreibbarkeit der Simulationen ermöglichte es, optimale Lösungswege und psychometrische Kennzahlen zu ermitteln. Zudem wurden Simulationen in ihrer Komplexität vergleichbar und beschreibbar. Damit wurden mehrere der größten Kritikpunkte an der bisherigen Forschung zum komplexen Problemlösen auf einen Schlag entkräftet. Entsprechend wurde die formalisierte Version der Schneiderwerkstatt schnell zur „Drosophila“ der Problemlöseforschung (Funke, 2010) und in einer Vielzahl von Publikationen verwendet (Fischer et al., 2011).

Es blieb aber das Problem des *Single Item Testing*, also der Verwendung von einzelnen, sehr zeitaufwendigen Aufgaben zur Erfassung der operativen Intelligenz. Dieses Problem löste Funke zusammen mit seinem Doktoranden Samuel Greiff, bei dem Funkes Ideen zu minimal-komplexen Systemen auf besonders fruchtbaren Boden fielen, durch eine elegante Minimierung komplexer Aufgaben auf die kleinstmögliche Anzahl von Variablen, die dennoch ein komplexes Problem beschreiben können (Funke & Greiff, 2017). Von diesen häufig nur aus 4 bis 6 Variablen bestehenden Problemen können problemlos bis zu 10 verschiedene Probleme in einem Test bearbeitet werden, ohne eine realistische Testzeit zu überschreiten. Diese Aufgabenminimierung ermöglicht zudem eine Trennung der Phase des Wissenserwerbs, in der das Problem erkundet wird, von der Phase der Wissensanwendung, in der das erworbene Wissen angewendet wird, um das System in einen bestimmten Zielzustand zu überführen (Greiff & Funke, 2017).

Auch wenn heute intensiv darüber diskutiert wird, ob dieser Weg der Fokussierung auf psychometrische Gütekriterien der richtige war (Funke selbst bezweifelt

dies; z. B. Funke, 2014), lässt sich nicht leugnen, dass die Idee der minimal komplexen Probleme wegweisend war.

Die internationalen Large-Scale-Assessments

Zusammen mit Joachim Funke hat Samuel Greiff den MicroDYN-Ansatz (Greiff & Funke, 2009), die bekannteste Umsetzung minimal komplexer Systeme, nun auf die große Bühne der internationalen Large-Scale-Assessments geführt. In der PISA-Studie 2012 bearbeiteten Schüler*innen aus 65 Ländern erstmals computerbasierte Aufgaben, darunter auch Aufgaben nach dem MicroDYN-Ansatz (OECD, 2014). Drei Jahre später wurden die Aufgabenstellungen um kooperative Probleme erweitert, wobei zwar nicht explizit der MicroDYN-Ansatz, aber die Idee mehrerer kleiner Probleme anstelle eines sehr realistischen Problems umgesetzt wurde (OECD, 2017).

Insbesondere die schon von Dörner geforderte Analyse von Verhaltensdaten konnte nun mit ungeahnten Stichprobenumfängen realisiert werden; allerdings bei wesentlich weniger komplexen Problemen. Als klar dominierende Strategie erwies sich die *Vary-One-Thing-At-A-Time* (VOTAT)-Strategie, also die isolierte Manipulation einzelner Variablen, um deren Bedeutung für das Gesamtsystem zu erkennen. Diese Strategie, die auch die Grundlage des wissenschaftlichen Experimentierens darstellt, ist inzwischen fast zu einem Synonym für das Lösen von Aufgaben nach dem MicroDYN-Ansatz geworden. In einer Vielzahl von Studien wurde untersucht, welche Problemlöser*innen VOTAT anwenden, welche verschiedenen Arten von VOTAT-Anwendungen es gibt und ob die Anwendung von VOTAT gefördert werden kann (z. B. Greiff et al., 2015).

Diese faktische Gleichsetzung der Anwendung der VOTAT-Strategie mit komplexer Problemlösung ist nicht ohne Kritik geblieben. Sie führte zu der bisher einzigen gemeinsamen Publikation von Dietrich Dörner und Joachim Funke, in der sie einen Common Ground darin finden, dass realistische Probleme nur schwer durch die Anwendung von VOTAT gelöst werden können (Dörner & Funke, 2017).

Als Beispiel führen sie eine Managerin an, bei der es kaum vorstellbar ist, dass sie ihre Probleme durch eine einfache Anwendung der VOTAT-Strategie lösen kann.

Meine Perspektive auf komplexes Problemlösen

Nach diesem kurzen Überblick möchte ich zum Abschluss meine aktuelle Perspektive auf das Gebiet der psychometrischen Forschung zum komplexen Problemlösen erläutern. Als Doktorand von Samuel Greiff repräsentiere ich die nächste Generation der Forschung zum komplexen Problemlösen.

In der Definition von Dörner unterscheiden sich komplexe Probleme von statischen Problemen durch vier Eigenschaften, auf die ich oben kurz eingegangen bin: Dynamik, Polytelie, Interdependenz und Intransparenz. Obwohl ich davon überzeugt bin, dass komplexe Probleme alle diese Eigenschaften aufweisen, glaube ich nicht, dass sie alle wirklich entscheidend sind, um komplexe Probleme von statischen Problemen zu unterscheiden. Nehmen wir ein klassisches Matrixproblem (z. B. Raven, 2003). Natürlich gibt es einen Zusammenhang zwischen den einzelnen Elementen (oder zwischen den Elementen und den entsprechenden Designregeln) dieser Aufgabe, und dieser Zusammenhang ist nicht direkt transparent, sondern muss durch Nachdenken erkannt werden. Wenn die Aufgabe mehrere verschiedene Symbole enthält, könnte man sogar von einer Polytelie sprechen, da mehrere Ziele erreicht werden müssen. Solche Aufgaben sind jedoch eindeutig nicht dynamisch. Es gibt keine Veränderung durch den Lösungsversuch oder im Laufe der Zeit.

Ich schlage daher eine andere Klassifikation des Problemlösens vor, die versucht, die beiden gegensätzlichen Positionen zur Messung komplexer Problemlösefähigkeiten zu vereinen. In meiner Klassifikation erfordert jedes Problemlösen, ob statisch oder komplex, das Erkennen von Zusammenhängen zwischen den Elementen des Problems (Reasoning). Dies zeigt sich bereits in der sehr hohen Validität konventioneller Intelligenzmessungen bei der Vorhersage realistischen Problemlösens. Um diese Zusammenhänge zu erkennen, ist VOTAT eine sehr effiziente Strategie. Bei manchen Problemen kommt jedoch ein dynamischer Aspekt hinzu,

der über das hinausgeht, was mit konventionellen Intelligenztests gemessen wird, und der mit VOTAT nicht erfasst werden kann. Der Umgang mit dieser Dynamik erfordert eigene Strategien, die eher dem entsprechen, was Dörner ursprünglich als *operative Intelligenz* bezeichnet hat. Unsystematisches oder dogmatisches Handeln ohne Reflexion, wie es Dörner in seinen ersten Experimenten beobachtete, ist nicht zielführend, wenn es z. B. zeitverzögerte Effekte in einem System gibt.

Intransparenz und Polytelie wären nach diesem Verständnis keine definitiven Eigenschaften eines Problems, sondern Elemente, die die Schwierigkeit eines Problems definieren. Ein komplexes Problem kann mehr oder weniger intransparent sein. Es wird nur umso schwieriger, je weniger offensichtlich die Zusammenhänge zwischen den Variablen sind (Stadler et al., 2019).

Wie geht es nun weiter?

Ein alter Witz beschreibt einen Mann, der nachts unter einer Straßenlaterne nach etwas sucht. Ein Passant will ihm helfen und fragt ihn, was er suche. Der Mann antwortet, er habe ein paar Meter weiter seinen Schlüssel verloren. Auf die Frage, warum er dann hier unter der Laterne suche, antwortet der Mann: „Weil hier Licht ist!“

Dieser Witz erinnert mich oft an die aktuelle Situation der Forschung zum komplexen Problemlösen. Dieses Forschungsgebiet hat in seiner über 50-jährigen Geschichte nichts von seiner Relevanz verloren. Die erschreckenden Vorhersagen des Club of Rome (Meadows et al., 1972), die 1972 die Initialzündung für die Erforschung komplexer Systeme und der menschlichen Unfähigkeit, deren Verhalten vorherzusagen, waren, sind heute Realität. Unsere Forschungsmethoden und Fragestellungen haben sich jedoch in Richtung abstrakterer, psychometrischer Fragestellungen verschoben. In Anlehnung an den Witz suchen wir die Antwort auf die ursprünglichen großen Fragen (den Schlüssel) nun einige Meter weiter, weil das Licht der psychometrischen Güte dort heller leuchtet. Damit will ich nicht sagen, dass wir wieder in die (psychometrische) Dunkelheit zurückkehren sollen, sondern dazu aufrufen, aus den vorhandenen Methoden eine Taschenlampe zu

bauen, mit der wir das Licht dorthin tragen können, wo wir den Schlüssel vermuten, um ihn endlich zu finden.

Ich hoffe sehr, dass uns Joachim Funke mit seinem großen Erfahrungsschatz beim Bau dieser Taschenlampe und beim Erkennen des Schlüssels unterstützen wird!

Referenzen

- Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*, 32(4), 290–308.
- Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (Hrsg.). (1983). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität*. Huber.
- Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8(1153), 1–11. <https://doi.org/10.3389/fpsyg.2017.01153>
- Fischer, A., Greiff, S., & Funke, J. (2011). The process of solving complex problems. *Journal of Problem Solving*, 4(1), 19–42. <http://dx.doi.org/10.7771/1932-6246.1118>
- Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning*, 7(1), 69–89. <https://doi.org/10.1080/13546780042000046>
- Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, 11(2), 133–142. <https://doi.org/10.1007/s10339-009-0345-0>
- Funke, J. (2014). Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Frontiers in Psychology*, 5(739), 1–3. <https://doi.org/10.3389/fpsyg.2014.00739>
- Funke, J., & Greiff, S. (2017). Dynamic problem solving: Multiple-item testing based on minimally complex systems. In D. Leutner, J. Fleischer, J. Grünkorn, & E. Klieme (Hrsg.), *Competence assessment in education* (S. 427–443). Springer International Publishing. https://doi.org/10.1007/978-3-319-50030-0_25
- Greiff, S., & Funke, J. (2009). Measuring complex problem solving: the MicroDYN approach. In F. Scheuermann & J. Björnsson (Hrsg.), *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (S. 157–163). Office for Official Publications of the European Communities. <https://doi.org/10.11588/heidok.00015502>

- Greiff, S. & Funke, J. (2017). Interactive problem solving: exploring the potential of minimal complex systems. In B. Csapó, & J. Funke (Hrsg.), *The nature of problem solving: Using research to inspire 21st century learning* (S. 93–105). OECD Publishing. <https://doi.org/10.1787/9789264273955-en>
- Greiff, S., Wüstenberg, S., & Avvisati, F. (2015). Computer-generated log-file analyses as a window into students' minds? A showcase study based on the PISA 2012 assessment of problem solving. *Computers & Education*, 91, 92–105. <https://doi.org/10.1016/j.compedu.2015.10.018>
- Meadows, D. H., Meadows, D. L., Randers, J., & Behrens, W. W. (1972). *The limits to growth*. Potomac Associates.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77–101. <https://doi.org/10.1037/0003-066X.51.2.77>
- OECD. (2014). *PISA 2012 results: Bd. 5. Creative Problem Solving: Students' Skills in Tackling Real-Life problems*. OECD Publishing.
- OECD. (2017). *PISA 2015 results: Bd. 5. Collaborative problem solving.*. OECD Publishing.
- Raven, J. (2003). *Raven progressive matrices*. Springer.
- Stadler, M., Niepel, C., & Greiff, S. (2019). Differentiating between static and complex problems: A theoretical framework and its empirical validation. *Intelligence*, 72, 1–12. <https://doi.org/10.1016/j.intell.2018.11.003>
- Süß, H.-M. (1999). Intelligenz und komplexes Problemlösen: Perspektiven für eine Kooperation zwischen differentiell-psychometrischer und kognitionspsychologischer Forschung. *Psychologische Rundschau*, 50(4), 220–228. <https://doi.org/10.1026//0033-3042.50.4.220>
- Toda, M. (1962). The design of a fungus-eater: A model of human behavior in an unsophisticated environment. *Behavioral Science*, 7(2), 164–183. <https://doi.org/10.1002/bs.3830070203>
- Wilhelm, O. (2009). Issues in computerized ability measurement: Getting out of the jingle and jangle jungle. In F. Scheuermann & J. Björnsson (Hrsg.), *The transition to computer-based assessment. New approaches to skills assessment and implications*

for large-scale testing (S. 145–150). Office for Official Publications of the European Communities. <https://op.europa.eu/s/zYfE>

Die Problemlösung ist noch lange nicht die Lösung des Problems

ANNETTE KÄMMERER

Ruprecht-Karls-Universität Heidelberg

Sebastian H. ist 23 Jahre alt. Er studiert im achten Semester Physik, die Masterarbeit steht an. Er kommt mit dieser Arbeit nicht voran, verschiebt alles von einem Tag auf den anderen, liegt oft unentschlossen im Bett, macht sich Vorwürfe. In die Therapie kommt er mit dem Anliegen, seine Blockade in Bezug auf die Masterarbeit zu überwinden. Sebastian ist ein kluger Kopf, in der Schule war er immer der Beste. Er hat noch einen jüngeren Bruder, der das Sorgenkind der Familie war. Der Vater ist in der Forstverwaltung tätig, die Mutter hat sich um den Haushalt und die beiden Söhne gekümmert. In der Familie, so wird im Gespräch deutlich, wurde wenig gesprochen, schon gar nicht über Gefühle. Jeder ging und geht seiner Wege. Zwar sind die Eltern stolz, dass Sebastian studiert, aber es besteht von Seiten der Eltern wenig Interesse an seinem Befinden. An der Universität hat Sebastian einige wenige soziale Kontakte, er trifft sich regelmäßig mit einigen Kommilitonen, meistens geht es allerdings um fachliche Aspekte. Konkurrenz spielt dabei eine große Rolle, es ist wichtig, wer gut oder sehr gut ist, alle anderen werden mit ziemlicher Verachtung angesehen.

Im therapeutischen Kontakt wird schnell deutlich, dass Sebastian sich mit einem analysierenden, an kognitiven Aspekten orientierten therapeutischen Vorgehen anfreundet. Seine aktuellen Schwierigkeiten als ein zu lösendes Problem anzusehen,

erleichtert ihn. Er beginnt, sich selbst differenzierter zu sehen, kann sich von den quälenden Selbstvorwürfen distanzieren, entdeckt seine schwachen und emotional unsicheren Seiten. Dieser Prozess der Selbsterkenntnis führt dazu, dass er seine Ziele zunehmend hinterfragt, nicht nur diejenigen, die sein Studium begleiten, sondern auch jene, die er sich für sein Leben gesetzt hat. Er erkennt, dass er in einer Atmosphäre von Leistungsansprüchen aufwuchs, gelernt hat, der unkomplizierte Sohn zu sein, der keine Schwierigkeiten hat oder diese, falls welche auftreten, meistert.

Die Problematik von Sebastian lässt sich leicht als ein komplexes Problem verstehen (siehe Funke 2003, S. 126–135):

- Vermeidung, Angst, Leistungsdruck, Unsicherheit, biografische Prägungen etc. machen das Ganze *komplex*, es sind vielfältige Gründe, die dazu beitragen, dass Sebastian mit seiner Masterarbeit nicht vorankommt.
- Die verschiedenen Facetten der Problematik sind miteinander *vernetzt*, sie verstärken sich gegenseitig, z. B. der Wunsch nach Anerkennung einerseits und das Erleben eigener Leistungseinbußen andererseits.
- Es herrscht eine gewisse *Dynamik*, nicht zuletzt durch die Tatsache, dass es Abgabefristen gibt, die zu einer Verschärfung der Problemkonstellation beitragen.
- Und nicht alle Aspekte der Problematik sind klar erkennbar, etwa der Einfluss biografischer Erfahrungen. Wir haben es demnach mit *Intransparenz* zu tun.
- Auch das letzte Kennzeichen komplexer Probleme, die *Polytelie*, ist im vorliegenden Fall zu berücksichtigen: Wie gelingt Sebastian eine gute *Work-Life-Balance*? Wie kann er die in der Therapie gewonnenen Erkenntnisse über sich und seine Lebensziele umsetzen?

Über alle diese verschiedenen Aspekte konnte Sebastian mühelos nachdenken und sie im therapeutischen Gespräch reflektieren und klären. Es waren rührende,

zum Teil traurige, aber auch erhellende, ehrliche Gespräche, die wir miteinander führten. Den Prozess der wachsenden Selbst(er)kenntnis dieses jungen Mannes zu begleiten, war eine befriedigende therapeutische Arbeit.

Aber die Masterarbeit war noch nicht geschrieben.

Damit kam Sebastian einfach nicht weiter. Wir verabredeten so ziemlich alles, was im Handwerkskasten bei solchen Problemkonstellationen empfohlen wird: Zeitstruktur, Strukturierung der Aufgaben, des Themas etc., von der Analyse der Befürchtungen ganz zu schweigen. Im direkten therapeutischen Gespräch konnte sich Sebastian auf alle diese Überlegungen gut einlassen, ging jedes Mal gestärkt und mit den besten Absichten nach Hause. Nur um mehr oder weniger erfolglos in der nächsten Stunde wieder vor mir zu sitzen. Er kam, um es mit den Worten von Heckhausen (z. B. 1989) zu sagen, einfach nicht über den *Rubikon*.

Trotz der analysierenden Durchdringung seiner Probleme blieb für den Patienten ein intentionales Problem hängen (Heckhausen, 1996). Die Problemlösetheorie suggeriert in einem solchen Fall, dass „immer dann, wenn sich bei der Umsetzung von Intentionen Schwierigkeiten ergeben, ... von einem zu lösenden Problem gesprochen werden kann“ (Funke, 2003, S. 99). Das bedeutet im Grunde, von vorne anzufangen und unter Einsatz bekannter Mittel, nämlich der kognitiven Analyse – also zum Beispiel dem Problemlösen –, der mangelnden Intention auf die Schliche zu kommen.

Aber ist das eine gute Idee? „Mehr desselben“ ist im Fall von Misserfolg kein guter Weg. Was fehlt also?

Innerhalb der psychotherapeutischen Forschung hat der Begriff des *Leidensdrucks* an dieser Stelle eine gewisse Relevanz. Dieser – so die Annahme – müsse nur groß genug sein und dann ergäben sich schon Veränderungen. In moderneren Ansätzen ist dieser Begriff durch den der *Veränderungsmotivation* ersetzt worden (z. B. Berking & Kowalsky, 2012). Andere Autoren sprechen von der Ausbildung von Annäherungszielen, die an die Stelle von Vermeidungszielen treten sollen (z. B. Grawe, 2004); Annäherungsziele seien mit positiven Emotionen verbunden und diese wiederum wirkten, verkürzt gesprochen, günstig auf die Veränderungsmotivation. Es wird dadurch deutlich, dass eine vorwiegend kognitive, analysierende

Durchdringung eines Sachverhalts – in diesem Fall die Prokrastination Sebastians bei der Masterarbeit – zu kurz greift und das emotionale Erleben eines Menschen viel stärker in Betracht gezogen werden muss. Und zwar nicht als eine Art *Beiprodukt* der reflektierenden Auseinandersetzung mit einem Problem, sondern als die eigentliche Variable im Veränderungsprozess.

Die empirische Evidenz der therapeutischen Nützlichkeit von Annäherungszielen ist in der psychotherapeutischen Prozessforschung in den letzten Jahren hinreichend belegt worden (z. B. Flückiger, 2021). Es ist für ein gelingendes seelisches Gleichgewicht eben nicht ausreichend, lediglich zu wissen, was man nicht möchte, denn daraus ergibt sich nicht ohne Weiteres, was stattdessen gewünscht wird. Positive Ziele, auch das ist Bestandteil der Problemlösetheorie, sind gar nicht so einfach zu formulieren, wie es vielleicht zunächst scheint. Denn (nicht nur) Patienten wissen vorwiegend, was sie nicht wollen, welche Gefühle und Erfahrungen nicht mehr das eigene Leben bestimmen sollen. Aber das Erkennen und Benennen von Wünschen, von Sehnsüchten, von Werthaltungen für das eigene Leben stellt gleichwohl eine Herausforderung dar.

Im Fall von Sebastian war, wie bereits erwähnt, die Masterarbeit vor allem mit Angst besetzt. Sie war ein wesentlicher Bestandteil seines Empfindens und sehr schnell auslösbar. Die Angst war facettenreich, Versagen, Enttäuschung, Verachtung, Scham, Schuld – das waren einige ihrer Auslöser. Wie also in einer solchen emotional belastenden Gemengelage zu positiven (Annäherungs-)Gefühlen und zu einer zielführenden, veränderungsorientierten Handlung kommen?

Im therapeutischen Arbeiten, und nicht nur da, bedarf es hierbei eines Schrittes, der aus meiner Sicht eines der größten Wunder der menschlichen Seele darstellt, nämlich der *Entscheidung* eines Menschen, einen neuen Weg zu gehen. Dieser Entschluss „Ich mache es jetzt anders“, ist immer wieder ein Schlüsselmoment in der therapeutischen Arbeit. Das Bild vom Rubikon, das Heckhausen gewählt hat, ist eigentlich wunderbar, denn es suggeriert ja, dass jemand wirklich einen neuen Weg einschlägt, mit allen Risiken, die dieser mit sich bringt. Das mag nicht so schwer sein, wenn das Selbstvertrauen groß ist und die zu gehenden Schritte

klar vor einem liegen. Aber je komplexer das Problem ist und je angeschlagener das Selbstvertrauen, desto schwerer wird es und desto mehr Mut ist notwendig.

Die psychologischen Modellierungen, die diesen Schritt zu beschreiben versuchen, sind vielfältig. Problemlösen ist eines dieser Modelle, die Analyse von Annäherungszielen oder Bedürfniskonstrukten ein anderes – und viele weitere Modelle ließen sich aufzählen. Ihnen ist gemeinsam, dass diese Entschlussbildung – dieser Ausdruck menschlicher Willensfreiheit – von verschiedenen Seiten beleuchtet wird, je nach psychologischer Subdisziplin. Letztlich, so meine Überzeugung, bleiben die psychologischen Modelle bestenfalls Näherungen an das Phänomen der Entschlussbildung und der Entscheidung, das eigene Leben zu ändern. Psychologische Forschung kann den Vorgang beschreiben und empirisch untersuchen. Aber das Geheimnis des inneren Schritts, dieses „Ja, ich werde, weil ich will“, können die psychologischen Theorien nur sehr vage abbilden.

Ich habe in meiner Arbeit als Psychotherapeutin viele mutige Menschen kennengelernt, Sebastian war einer davon. Schlussendlich hat er die Kurve gekriegt. Das Problemlösen hat uns geholfen, sein Problem zu analysieren und zu verstehen. Den Entschluss, sein Herz in die Hand zu nehmen und einen anderen Weg zu gehen, den Mut zu haben, sich mit sich selbst zu konfrontieren, der Angst ins Auge zu schauen und sie zu lindern, konnte ihm die kognitive Analyse allein allerdings nur mäßig erleichtern.

Referenzen

- Berking, M., & Kowalsky, J. (2012). Therapiemotivation. In M. Berking & W. Rief (Hrsg.), *Klinische Psychologie und Psychotherapie für Bachelor: Bd. 2. Therapieverfahren. Lesen, Hören, Lernen im Web.* (S. 13–22). Springer. https://doi.org/10.1007/978-3-642-25523-6_2
- Flückiger, C. (2021). Basale Wirkmodelle in der Psychotherapie. Wer und was macht Psychotherapie wirksam? *Psychotherapeut*, 66(1), 73–82. <https://doi.org/10.1007/s00278-020-00478-y>

- Funke, J. (2003). *Problemlösendes Denken*. Kohlhammer. <https://doi.org/10.17433/978-3-17-022830-6>
- Grawe, K. (2004). *Neuropsychotherapie*. Hogrefe.
- Heckhausen, H. (1989). *Motivation und Handeln*. Springer. <https://doi.org/10.1007/978-3-662-53927-9>
- Heckhausen, H. (1996). Intentionsgelitetes Handeln und seine Fehler. In J. Kuhl & H. Heckhausen (Hrsg.), *Enzyklopädie der Psychologie: Bd. 4. Motivation, Volition und Handlung* (S. 817–845). Hogrefe.

Die Universität ist kein Wirtschaftsunternehmen!

BEN ROEG UND PETA PANTA

Köln und Heidelberg

Blog 2233

Das neue Semester hat auch in der Universitätsspitze umwälzende Veränderungen zu verzeichnen, in erster Linie im Kanzleramt. Die neue Kanzlerin ist eine Quereinsteigerin aus der Wirtschaft, die sich bereits einen Namen bei der Sanierung maroder Unternehmen gemacht hat. Unser naturwissenschaftlicher Rektor hat bei ihrer Einführung begierig die Gelegenheit ergriffen und mit seinem unvergleichlichen Humor sprachwissenschaftliche Bildung demonstriert: „Honni soit qui mal y pense!“ Natürlich bedeute das einmütige Votum des Senats für diese personelle Erneuerung keineswegs, dass irgendjemand unsere ehrwürdige Alma Mater für ein marodes Unternehmen halte! In keinster Weise! Im Gegenteil: Es gehe einzig und allein darum, unserer Universität im Konzert der exzellenten Wissenschaftsstandorte eine noch herausragendere Position zu verschaffen, indem das wirtschaftliche Fundament für eine auch international sichtbare Ausnahmestellung geschaffen wird. Das Ziel ist die Bündelung aller Kräfte, auf dass sich aus dem Breitensport der alle Fächer anbietenden Universität die Rekorde global anerkannter spezieller Fachdisziplinen entwickeln können. – Und wer möchte

da widersprechen, wenn Wirtschaften und Verwalten in der Hochschule noch effizienter in den Dienst der Wissenschaft gestellt werden?

Also weht jetzt ein frischer Wind durch die Flure der Erkenntnis. Das Bemühen um einen neuen Stil ist erkennbar, nicht zuletzt durch das Interview mit der studentischen Campus-Zeitschrift. In dem es ersichtlich nicht nur um die wirtschaftlichen Probleme und Aussichten der Universitätsentwicklung gehen sollte, sondern auch um den notwendigen *Human Touch*, mit dem man heute um Zustimmung für alle offenbaren und noch mehr die intransparenten Ziele und Wege werben muss. Also hat unsere Kanzlerin auch mit der emotionalen Antwort überrascht, wie stimmungsvoll der „ökonomische Gottesdienst“ zur Semestereröffnung war. (Ein Schelm, wer Böses dabei denkt.) Denn im Prinzip hat sie ja recht. Wenn die Ökumene endlich zum Erfolg käme und die (Wieder-)Vereinigung der christlichen Kirchen erreichen würde, wäre das auf jeden Fall ein ungeheurer ökonomischer Gewinn. Welch eine Verschwendung, dass in vielen Städten katholische und evangelische Gotteshäuser nebeneinander bestehen, das heißt eben nicht zuletzt auch: unterhalten werden müssen. Doch bekanntlich schlagen Personalkosten am meisten zu Buche, in diesem Fall etwa katholische Pastoren und evangelische Pfarrerrinnen. All diese Posten auf eine Linie zu bringen, wäre eine unbezahlbare Verschlangung nicht nur finanzieller, sondern vor allem auch spiritueller Synergie! Wieso also sollte das nicht genauso eines der hehrsten Ziele jeder Universitätsentwicklung sein: die Verschlangung der Geistesgeschichte?

Blog 2244

In der Tat, die Hörsäle und Seminarräume erstrahlen neuerdings in einem beeindruckenden Glanz. Nicht nur neues Mobiliar, sondern auch eine Technik vom Feinsten: Beamer selbstverständlich, aber ebenfalls eine hervorragende Mikro-Lautsprecher-Anlage. Alles ein Erfolg des Sponsoring, das unsere Kanzlerin – wie sagt man heute – eingeworben hat! Das ist doch einmal eine fruchtbare Verbindung mit der Wirtschaft, dass die Unternehmen mit ihrer Unterstützung der Wissenschaft werben. Zumindest bei denjenigen, die jetzt zu einer Lehrveranstaltung nicht mehr

in den Hörsaal III strömen, sondern in den Firma-XYZ-Hörsaal! Selbstverständlich sind auch alle Institute gehalten, ab sofort ihre Vorträge für die interessierte Öffentlichkeit mit den korrekten, vollständigen Orts- und Raumnamen anzukündigen. Was zudem in jedem Eingangsbereich zur Orientierung überdimensionierte Tafeln mit Seminar- und Hörsaalnamen erfordert, bei denen die kurze Stockwerkanzeige fast untergeht. Aber die Firmenwerbung steht dann nicht zuletzt genauso in der Zeitung: Öffentlicher Vortrag „Vorhofflimmern, die neue Volkskrankheit“ am XX.YY. in der Universität, Unternehmens-ABC-Hörsaal, um 18 Uhr.

Da steht durchaus zu erwarten, dass sich die übernommenen Renovierungskosten für die beteiligten Unternehmen *rechnen*. Insofern hat die Kanzlerin den Firmen wohl nicht zu viel versprochen, vorausgesetzt natürlich, dass die Institute der Universität bei dieser schönen Win-Win-Strategie engagiert mitmachen. Einziger Schönheitsfehler ist allerdings, dass es keinerlei inhaltliche Übereinstimmung zwischen Raum-Namen und darin präsentierten Erkenntnissen gibt. Sondern im Gegenteil nicht selten sogar einen Gegensatz: „Gesunde Ernährung“ im Petro-Chemie-Hörsaal oder „Eigentum verpflichtet“ im Sparkassen-Seminarraum... Hier gibt es also noch Verbesserungsmöglichkeiten. Deren eklatanteste ist merkwürdigerweise vom Kanzleramt übersehen oder (vielleicht sogar in konventioneller Prüderie?) intentional vermieden worden: Die Toiletten sind immer noch in einem bejammernswerten Zustand, obwohl ihre Renovierung unter Beistand eines auch noch regional ansässigen Unternehmens (<https://www.hakle.de/>) längst überfällig wäre: Hakle-Terminals!

Blog 2255

Das Kanzleramt überrascht doch immer wieder mit unerwarteten Nachrichten. Heute die Pressemitteilung, dass die Universität das – natürlich äußerst lukrative – Angebot einer bekannten Unternehmensberatung angenommen hat, die Verwaltungsstrukturen der Hochschule auf allfällige Verbesserungsmöglichkeiten zu überprüfen. Ist das nun das Eingeständnis der Kanzlerin, dass sie doch nicht so viel von Wirtschaft versteht, wie ihr als Ruf vorausgeeilt ist? Wohl kaum! Aber

wozu dann diese fulminante Geldausgabe, um sich von Spezialisten opaker Expertise beraten zu lassen? Kennt denn niemand in der Universitätsspitze die schöne Geschichte vom Schäfer und seinem Überraschungsgast?

Also: der Schäfer mit seiner Herde auf einer Voralpen-Alm. In einer Staubwolke kommt ein Jeep angebraust, dem ein junger Mann im Brioni-Anzug, mit Hermès-Krawatte etc. entsteigt. „Guter Mann“, sagt er, „wenn ich Ihnen genau sagen kann, wie viele Schafe Sie hier haben, krieg ich dann eines?“ Der Schäfer ist friedlich gestimmt und nickt. Der junge Mann packt sein Handy aus, schließt es an den Laptop an, geht auf einige NASA-Seiten, öffnet eine Reihe von Excel-Dateien und verkündet nach zehn Minuten: „Sie haben genau 1.576 Schafe!“ „Stimmt“, sagt der Schäfer, „nehmen Sie sich eins.“ Der junge Mann packt ein Tier in seinen Kofferraum, schwingt sich hinters Steuer und will losbrausen. Da fragt ihn der Schäfer: „Und wenn ich errate, welchen Beruf Sie haben, bekomme ich dann mein Tier zurück?“ „Okay“, ist die Antwort. „Sie sind Unternehmensberater!“ „Richtig“, gibt der junge Mann überrascht zu, „wie sind Sie darauf gekommen?“ „Ganz einfach. Erstens kommen Sie hierhin, ohne dass ich Sie benötigt habe. Zweitens wollen Sie eine Bezahlung dafür, dass Sie mir sagen, was ich eh schon weiß. Und drittens haben Sie keine blasse Ahnung von dem, was ich hier tue. Also, geben Sie mir jetzt meinen Hund zurück?“

Eben: QED!

Blog 2266

Siehe da, es geht doch. Mehr als ein Jahrzehnt hat das Institut vergeblich auf zusätzliche Räumlichkeiten gehofft, Jahr für Jahr gut begründete Anträge verfasst... Aber selbst Drittmittelinwerbungen von erheblichem Ausmaß (Forscherguppe und dergleichen) haben die Universitätsspitze nicht bewegen können; vor allem das Kanzleramt gefiel sich in seiner Rolle als erratischer Block der Standhaftigkeit.

Doch jetzt auf einmal schlägt das Pendel ins absolute Gegenteil um: Es gibt einen ganz neuen zusätzlichen Standort, den das Institut personell (noch) gar nicht vollständig besetzen kann. Aber, so das äußerst wohlwollende Argument der Kanzlerin, das neue Psychotherapeutengesetz erfordert absehbarer Weise eine erkleckliche Menge an weiteren Praxis- und Supervisionsräumen, um die Anforderung zu erfüllen, dass die Studierenden der Klinischen Psychologie schon innerhalb des Master-Studiums die Bedingungen für eine Approbation erfüllen!

Und wie recht sie hat! Die Möglichkeit, die Approbation bereits im Studium zu erreichen, vergrößert die Attraktivität der Klinischen Psychologie für die Studierenden noch einmal um ein Vielfaches. Es wird einen übergroßen Masterstudiengang Klinische Psychologie/Psychotherapie geben, alle anderen Schwerpunktsetzungen von Pädagogischer Psychologie über Arbeits-, Betriebs-, Organisations- bis zur Verkehrs-, Rechtspsychologie etc. werden zu einer Restkategorie eingedampft, für die das Hauptproblem darin besteht, eine so generelle Benennung zu finden, dass alles eingemeindet, nichts ausgeschlossen wird. Und auch im Bachelorstudiengang werden die sog. Grundlagenfächer (von Allgemeiner über Sozial- und Persönlichkeits- bis zur Entwicklungspsychologie) nur mehr zu Rahmenkonturen, die das Tor zu möglichst frühzeitiger Praxis öffnen (sollen). Praxis aber erfordert möglichst kleine Lern- und Übungsgruppen mit einer entsprechenden Vermehrung von Lehrkräften und Raumbedarf. Die Psychologie ist am Höhepunkt ihrer Praxisrelevanz und damit auch ihres Ausbaus in der Universität angekommen. Berechtigter Weise! Unsere Alma Mater kann sich glücklich schätzen, dass diese Entwicklung des Fachs auf eine Kanzlerin trifft, die Chancen und Notwendigkeiten einer solchen gesellschaftlichen Relevanz unmittelbar zu erkennen und mit konstruktiven Haushaltsentscheidungen (im Personaltableau wie in der Raumplanung) zu beantworten in der Lage ist.

Selbstverständlich gibt es auch in diesem Feld der Universitätsentwicklung – wie überall – zögerliche, abwehrende Bedenkenträger. Deren lautstärkster ist ein Kollege namens Clemens Knobloch mit seinem Elaborat (von 2010): *Wir sind doch nicht blöd. Die unternehmerische Hochschule*. Darin beklagt er mit dramatischem Pathos die „Übernahme der Universitäten durch die Wirtschaft“ (Knobloch, 2010,

S. 227). Den Hauptindikator bzw. die Folge davon sieht er in der Tendenz, dass nun „allenthalben die Praxis (oder was dafür gilt) in die Hochschulausbildungen hinein“ drängt (Knobloch, 2010, S. 229). Dadurch bleibt für ihn das „allgemeine Grundrecht auf Bildung ... an der unternehmerischen Universität auf der Strecke“ (Knobloch, 2010, S. 233). An deren Stelle träten „Studiengänge, deren Inhalte sich vor den Kunden und Abnehmern, vor den täglich wechselnden Anforderungen des Arbeitsmarktes zu rechtfertigen haben“ (Knobloch, 2010, S. 242). An die Stelle der ‚Einheit von Forschung und Lehre‘ trete damit die ‚von Wissen und Gewinn‘ (Knobloch, 2010, S. 242). Und wie es sich für einen Mahner in der Wüste gehört, sieht er letztlich das Weltende heraufziehen: „Als zeitgeistiger Themenpark hat die Uni keine Überlebenschance“ (Knobloch, 2010, S. 252).

Glücklicherweise steht nicht zu befürchten, dass sich unsere Kanzlerin von solchen überzogenen Kassandrarufern wird beirren lassen. Für den Kollegen Knobloch wäre die Entwicklung des Fachs Psychologie, das nach dem Psychotherapeutengesetz in absehbarer Zeit zu etwa 90 % (der Studierenden wie Lehrinhalte) aus Klinischer Psychologie bestehen wird, sicher ein paradigmatisches Beispiel für die Übernahme der Universität durch die Wirtschaft. Aber es handelt sich letztlich vor allem um die Übernahme von gesellschaftlicher Verantwortung! Denn es besteht nun einmal – gerade in unserer komplexen, globalisierten Welt – dieser eminente Bedarf an psychotherapeutischer Versorgung, für die eine Konzentration des Fachs Psychologie auf die Klinische Psychologie die konstruktive Antwort darstellt. In der Tat werden dadurch zum Beispiel Ausbildungsinhalte der Methodenlehre heruntergefahren, die es aber in der bisherigen Ausführlichkeit eigentlich sowieso nur für eine wissenschaftliche Karriere braucht. Selbstverständlich müssen gerade praktizierende klinische Psychologen/innen im Sinne einer evidenzbasierten Medizin neue und neueste Forschungsergebnisse rezipieren können. Aber dazu muss man nicht selbst in einem experimentellen Praktikum empirische Untersuchungen durchgeführt haben. Man lese dazu schon bei Lessing im 17. Literaturbrief nach: Wer Romane kritisieren will, muss nicht selbst Romane schreiben können! Forschungskompetenz zu entwickeln, ist sicher eine sinn- und wertvolle Aufgabe, aber nur innerhalb von Promotionsstudienphasen...

Ähnliches ist dem Argument entgegenzuhalten, dass die differenzierte Kenntnis der angesprochenen Grundlagenfächer nötig sei, um eine rationale, abwägende Haltung gegenüber der Gesellschaft wie auch der eigenen Disziplin zu entwickeln. Es ist schon richtig, dass ein universitäres Studium ganz grundsätzlich zum kritischen Hinterfragen sowohl gesellschaftlicher Selbstverständlichkeiten als auch eigener Einstellungen befähigen soll. Aber das ist auch ohne exzessive Reflexionsangebote zu erreichen! Der Schlüssel zum Erfolg lautet hier: Learning by doing! Das ist gerade der Zauberschlüssel, der in der frühzeitigen Einübung von praktischen Kompetenzen ohne übertriebene Grundlagenorientierung liegt. Die Reflexion stellt sich anhand der positiven wie negativen Praxiserfahrungen von selbst ein – zumal sie von den vorgesehenen Supervisions-Elementen essenziell gefördert wird. Und die Voraussetzungen dafür werden ja gerade durch Ausweitung des Lehrkörpers und der Raumausstattung geschaffen!

Nein, es bleibt dabei: Das goldene Zeitalter der Psychologie hat gerade begonnen! Das Fach kommt seiner herausragenden gesellschaftlichen Verantwortung nach und wird dabei von einem engagierten Kanzleramt nachdrücklich und nachhaltig unterstützt. Gerade an unserer Universität, die damit als weithin sichtbarer Leuchtturm des universitären Fortschritts heraussticht!

(Literatur: Clemens Knobloch [2010]. *Wir sind doch nicht blöd. Die unternehmerische Hochschule*. Westfälisches Dampfboot.)

Blog 2277

Auch in der Digitalisierung soll unsere Alma Mater mit gutem, nein bestem Beispiel vorangehen. Dazu gehört nicht zuletzt die Doktoranden-Karte, über die alle möglichen Funktionen (von Rückmeldung bis Abbuchen des Semesterbeitrags, von Kopierguthaben bis Nahverkehrsticket, von UB-Ausleihe bis Mensa-Essen) erledigt werden können. Die Universitätsspitze hat hier ihre langjährige Verbindung zu einer international tätigen (ausländischen) Bank aufs Günstigste nutzen können, um den Studierenden ein optimales Angebot zu offerieren. Dass damit der Einfachheit halber auch der Erwerb einer Giro- bzw. Kreditkarte verbunden sein

wird, stört nur aus der Zeit gefallene Bedenkenträger. Denn dass die Studierenden auf diese Weise ggf. in die Kundenbindung zu einer bestimmten Bank gezwungen würden, ist heute kaum mehr realistisch anzunehmen, wo der Einzelne sowie so über eine Vielzahl an (Kredit-)Karten verfügen muss, um im kapitalistischen Wirtschaftskreislauf die notwendige Bonität nachzuweisen.

Solcherart Einwände müssen die Universitätsspitze also im Prinzip keineswegs beunruhigen. Aber überraschenderweise hat an dieser Stelle kürzlich die Staatsanwaltschaft die Rolle des Schelms eingenommen. Und das nur, weil die erwähnte langjährige Verbindung etliche Großspenden dieser Bank an die Universität mit umfasst, sodass der Verdacht aufgekommen ist, die Ausschreibung für die Entwicklung der Karte könne nicht korrekt erfolgt, sondern auf den jahrelangen Großspender zugeschnitten worden sein. Doch die unbestrittene Expertise unserer Kanzlerin auch in fiskalischen Verwaltungsabläufen gibt die größtmögliche Sicherheit, dass alles zumindest nach Aktenlage schön säuberlich getrennt worden ist, dass also auf keinen Fall die Ausweitung der staatsanwaltlichen Ermittlungen zu einem Schelmenroman droht!

Blog 2288

Da ich in dieser Angelegenheit eindeutig höchst engagierte Partei bin, versuche ich die mir größtmögliche Objektivität zu realisieren, indem ich das einschlägige Schreiben des Rektorats an den Dekan unserer Fakultät ohne Kommentar lediglich zitiere:

Spectabilis!

Sie wissen, wie sehr wir es begrüßen, wenn Fakultäten ihre herausragenden Kollegen für eine Senior Professur nach Abschluss der aktiven Dienstzeit vorschlagen. Es ist im ureigensten Interesse unserer Universität, dass die fähigsten Köpfe der Scientific Community auch über die beamtenrechtlich vorgeschriebene Altersgrenze hinaus in der Wissenschaft, zumal in der international sichtbaren Forschung, gehalten werden. Die Universitätsspitze hat daher keine Mühen gescheut, für alle Senior Professuren eine Art Grundausstattung (nicht nur räumlicher Art) bereitzustellen, auf dass die Koryphäen

ihrer jeweiligen Fächer, die für unsere Universität zu gewinnen in den letzten Jahrzehnten ein großer Erfolg war, weiter das exzellente Bild unserer Alma Mater mitbestimmen. Deshalb war und ist es uns ein Herzensanliegen, jeden Antrag auf Verleihung einer Senior Professur mit äußerstem Wohlwollen zu prüfen.

So ist es auch mit Ihrem Antrag vom XX.YY.ZZZZ geschehen, für den wir uns noch einmal ausdrücklich bedanken. Allerdings ist selbstverständlich bei jedem Antrag auch zu berücksichtigen, dass trotz aller finanziellen Bemühungen die Mittel für die erwähnte Grundausstattung der Senior Professuren beschränkt sind und bleiben müssen; denn dadurch darf auf keinen Fall eine Beeinträchtigung der regulären Forschung und Lehre in Kauf genommen werden. Deshalb können Senior Professuren wirklich nur an die Crème de la Crème der international anerkannten Forscher/innen vergeben werden. Und hier ist bekanntlich die Konkurrenz gerade an unserer Exzellenz-Universität besonders groß! Ein Faktum, das bei aller Begrenzung für die Vergabe von Senior Professuren das Rektorat jedoch verständlicher Weise vor allem mit Stolz erfüllt. Das ist auch der Grund, weswegen die Ablehnung von Anträgen auf eine Senior Professur die Universitätsspitze zwar einerseits schmerzt, andererseits jedoch mit dem Gefühl einer unvermeidbaren Folge des herausragenden Forschungsprofils unserer Universität ausgesprochen wird.

Das gilt nun – leider – auch für den von Ihnen eingereichten Antrag. Die Verdienste des in Frage stehenden Kollegen um unsere Universität sind völlig unbestritten, insbesondere was die Stärkung der Corporate Identity angeht. Wir haben sein ehrenamtliches Engagement in den verschiedensten Funktionen mit größter Bewunderung verfolgt; es seien pars pro toto nur die Freunde der Universität, die verschiedenen Stiftungsräte und der Aufbau eines Alumni-Netzwerks in seinem Fach und darüber hinaus erwähnt. Und wir hoffen inständig, dass er dieses Engagement auch nach seiner aktiven Dienstzeit fortzuführen gedenkt. Denn die Universität ist auf solche sozial kompetenten Netzwerker unbedingt angewiesen. Es ist vermutlich sogar höchstwahrscheinlich, dass dadurch eine erkleckliche Anzahl von Sponsoren gewonnen werden konnten. Wir haben das allerdings im Einzelnen nicht

in die Überprüfung und Entscheidung einbezogen, weil es sowieso kein entscheidendes Gewicht haben darf. Denn:

Die Universität ist kein Wirtschaftsunternehmen!

gez.: Magnifizenz [schwarz]

Teil III
Originalia

Problemlösen und implizites Lernen

HILDE HAIDER

Universität zu Köln

Abstract. Ziel des Beitrages ist es, eine Verbindung zu ziehen zwischen dem Thema Problemlösen auf der einen Seite und dem Thema implizites Lernen auf der anderen Seite. Über einen kurzen historischen Abriss der Problemlöseforschung wird herausgearbeitet, wie sich hieraus der Forschungsbereich implizites Lernen entwickelt hat. Im Weiteren werden Unterschiede und Gemeinsamkeiten beider Forschungsbereiche aufgezeigt.

Im Jahr 1985 begann ich an der Universität der Bundeswehr in Hamburg bei Rainer Kluwe zu promovieren. Das Thema war damals komplexes Problemlösen, ein seit den 70er-Jahren in der deutschen Psychologie sehr stark beforschtes Gebiet, das von Dietrich Dörner (1981) und Joachim Funke (1986) nachhaltig geprägt wurde. Das große Interesse am komplexen Problemlösen ergab sich einerseits aus der Fortführung der deutschen Denkpsychologie (Wundt, 1907) und andererseits aus drängenden angewandten Fragen: Wie kann man in einer vernetzten Umwelt Menschen darin unterstützen, die komplexen gesellschaftlichen Herausforderungen, wie z. B. Umweltverschmutzung, Waldsterben oder Ressourcenverknappung, zu bewältigen (Vester, 1974)?

Bei der Umsetzung offenbarten sich forschungspraktische Probleme. So zeigten die Untersuchungen häufiger, dass schlechte Problemlöseleistungen mit guten verbalen Beschreibungen des Problems verbunden waren, während eine gute Performanz beim Problemlösen mit einer geringen Fähigkeit einherging, die Struktur

des komplexen Systems zu beschreiben (Kluwe et al., 1990). Auch Versuche, selbst-reflexive Kompetenzen zu induzieren, blieben erfolglos (Putz-Osterloh, 1985). Komplexe Fehleranalysen, die Dörner und Kollegen (1983; 1988) anhand von Computersimulationen durchführten, erlaubten wenig präzise Vorhersagen, warum spezifische Fehler bei der Kontrolle komplexer Probleme auftraten (Funke, 1992). Insgesamt wurde deutlich, dass vermutlich noch zu wenig verstanden wurde, welche kognitiven Prozesse komplexem Problemlösen zugrunde lagen und wie diese gefördert werden können, um Menschen bei der Lösung komplexer Probleme zu unterstützen. In der Folge führte dies dazu, dass weniger komplexe Systeme (finite Automaten) für die Untersuchungen genutzt wurden, die einfacher zu formalisieren waren und damit klarere experimentelle Manipulationen ermöglichten (Buchner, 1995; Funke, 1986).

Auch der Einfluss von Donald Broadbents Arbeiten veränderte den Fokus der komplexen Problemlösungsforschung (Berry & Broadbent, 1984, 1988; Broadbent et al., 1986). Ähnlich wie Funke verwendeten die Autoren einfachere Probleme, die auf linearen Gleichungssystemen basierten. Diese Arbeiten offenbarten ein interessantes Phänomen: Menschen schienen die Systeme auf der Grundlage implizit erworbenen Wissens zu kontrollieren. Sie waren in der Lage, ein komplexes System optimal zu steuern (potenziell implizites Lernen), konnten aber gleichzeitig Fragen zur internen Struktur dieses Systems (potenziell explizites Wissen) nicht beantworten. Darüber hinaus zeigten Hayes und Broadbent (1988), dass die Kontrolle des Systems und strukturelles Wissen über das System durch unterschiedliche experimentelle Manipulationen beeinflusst wurden.

Diese Unterscheidung zwischen implizitem und explizitem Lernen (oder Wissen) beim komplexen Problemlösen löste jedoch schnell Kritik aus. Einerseits könnte eine solche zu beobachtende Dissoziation zwischen der Kontrollleistung und der Fähigkeit, Fragen zur Struktur des Systems zu beantworten, tatsächlich darauf zurückzuführen sein, dass die Kontrolle implizit erlernt wurde. Andererseits könnte diese jedoch auch dadurch erklärt werden, dass die Systemsteuerung auf einer linearen Optimierung des Systems basiert, die ohne Kenntnis der genauen Struktur des zugrunde liegenden Systems genutzt werden kann (Buchner, 1993, 1995;

Buchner & Funke, 1993; Haider, 1992). Inspiriert von diesen Debatten, begann ich mich in meiner Forschung auf die Frage zu konzentrieren, ob es möglich sei, zu lernen, ohne sich des Lernprozesses selbst sowie der gelernten Inhalte bewusst zu sein (implizites Lernen).

Implizites Lernen geht unter anderem auf die Arbeit von Reber (1967) zurück. Er entwickelte die sogenannte künstliche Grammatikaufgabe. Während der Lernphase werden die Versuchspersonen aufgefordert, Buchstabenfolgen zu beobachten, die aus einer komplexen zugrunde liegenden Grammatik abgeleitet wurden. Danach erhalten sie in der Testphase neue, der Grammatik entsprechende oder nicht entsprechende Buchstabenfolgen und sollen beurteilen, ob diese grammatikalisch korrekt sind. Reber und Kollegen (Reber, 1967; Reber & Lewis, 1977) beobachteten, dass die Versuchspersonen die Buchstabenfolgen besser beurteilen konnten, als per Zufall zu erwarten. Gleichzeitig waren sie aber nicht in der Lage, die Grammatik verbal zu beschreiben. Ähnlich wie bei Broadbents Arbeiten kam Kritik an diesen Befunden auf. Die Leistung in diesen künstlichen Grammatikaufgaben konnte entweder durch die Annahme erklärt werden, dass die Versuchspersonen implizit abstraktes Regelwissen über die künstliche Grammatik erworben hatten, oder – deutlich einfacher – durch das explizite Erlernen von Fragmenten der Buchstabenfolgen (Perruchet & Pacteau, 1990; Shanks & St. John, 1994). Dennoch wurde implizites Lernen zu einem der wesentlichen Themen in der kognitiven Psychologie.

Um diese Probleme zu umgehen, begannen Peter Frensch und ich eine etwas andere Frage an die implizite Lernforschung zu stellen. Die zugrunde liegende Beobachtung war, dass in fast jedem impliziten Lernexperiment einige wenige Versuchspersonen explizites Wissen über die zugrunde liegende Regularität erwerben. Daher fragten wir, wie sich explizites Wissen in einem impliziten Lernexperiment entwickeln könnte, wenn Versuchspersonen nicht über die Existenz einer in die Aufgabe eingebauten Regularität informiert werden (Frensch et al., 2003). Im Unterschied zu anderen Forschern verglichen wir nicht eine implizite Lernsituation mit einer Situation, in der die Versuchspersonen aufgefordert werden, nach einer zugrunde liegenden Gesetzmäßigkeit zu suchen (explizites Lernen; Dienes

& Berry, 1997), sondern nutzten leicht zu erkennende Regularitäten und testeten nach dem Experiment, ob die Versuchspersonen in der Lage waren, die Regularität vollständig verbal zu berichten.

Die Befunde zeigten, dass die Mehrheit der Versuchspersonen, die in der Lage waren, die gesamte Regularität in einem postexperimentellen Wissenstest zu benennen, irgendwann während des Trainings eine abrupt auftretende starke Abnahme der Reaktionszeiten aufwiesen. Wir vermuteten, dass diese Reaktionszeitabnahme den Moment widerspiegeln könnte, in dem die Versuchspersonen von einer stimulus- zu einer top-down-gesteuerten Kontrolle wechselten (Tubau et al., 2007). Erste Unterstützung dieser Annahme kam von Neuroimaging-Daten (Rose et al., 2010; Wessel et al., 2012). Diese ließen erkennen, dass, sobald der steile Abfall der Reaktionszeiten auftrat, die EEG-Signale eine Kopplung in der Gammaband-Aktivität indizierten. Darüber hinaus zeigte das BOLD-Signal in einer fMRT-Studie eine erhöhte Aktivität im medialen und ventrolateralen präfrontalen Kortex sowie im ventralen Striatum, kurz bevor die steile Abnahme der Reaktionszeiten sichtbar wurde.

Zur weiteren Prüfung drehten wir die Analysen um und nutzten die abrupt auftretende Verkürzung der Reaktionszeit, um die nach dem Training erfasste Menge an explizit erworbenem Wissen vorherzusagen (Haider et al., 2011). Gleichzeitig präsentierten wir im Training stroopartiges Aufgabenmaterial, um herauszufinden, ob der Reaktionszeitunterschied zwischen kongruenten und inkongruenten Stroop-Stimuli verschwinden würde, sobald sich die plötzliche Abnahme der Reaktionszeit bei einer Versuchsperson zeigt. In beiden Studien ergab sich bei ca. 50 % der Versuchspersonen eine abrupte Reaktionszeitabnahme während des Trainings. Von diesen konnte die Mehrheit die gesamte Regularität benennen und auch der Unterschied zwischen kongruenten und inkongruenten Stroop-Stimuli verschwand mit dem Moment des Auftretens der abrupten Reaktionszeitabnahme. Hingegen blieb bei den Versuchspersonen, die keine plötzliche Abnahme der Reaktionszeit zeigten, der Kongruenzeffekt bis zum Ende des Trainings bestehen und die meisten konnten nur Teile der gesamten Regularität verbalisieren. Diese Ergebnisse sprechen folglich dafür, dass (a) sich implizite und explizite Prozesse unterscheiden

und (b) explizites Wissen in einer impliziten Lernsituation ähnlich einer *Einsicht* innerhalb eines kurzen Zeitfensters entsteht.

Insgesamt veranschaulicht diese Forschung, dass implizite Lernexperimente geeignet sein können, um einen wesentlichen und interessanten Teilaspekt der Problemlöseforschung zu untersuchen, nämlich wie Menschen zu einer Einsicht gelangen und eine Umstrukturierung der Repräsentation einer Situation vornehmen. Es wäre sicher interessant, solche Ansätze zu nutzen, um die Prozesse, die komplexem Problemlösen zugrunde liegen, genauer zu ergründen.

Referenzen

- Berry, D. C., & Broadbent, D. E. (1984). On the relationship between task performance and associated verbalizable knowledge. *The Quarterly Journal of Experimental Psychology*, 36, 209–231. <https://doi.org/10.1080/14640748408402156>
- Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, 79, 251–272. <https://doi.org/10.1111/j.2044-8295.1988.tb02286.x>
- Broadbent, D. E., Fitzgerald, P., & Broadbent, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, 77, 33–50. <https://doi.org/10.1111/j.2044-8295.1986.tb01979.x>
- Buchner, A. (1993). *Implizites Lernen. Probleme und Perspektiven*. Psychologie Verlags Union.
- Buchner, A. (1995). Basic topics and approaches to the study of complex problem solving. In P. A. Frensch & J. Funke (Hrsg.), *Complex problem solving: The European perspective* (S. 27–63). Psychology Press. <https://doi.org/10.4324/9781315806723>
- Buchner, A., & Funke, J. (1993). Finite-state automata: Dynamic task environments in problem solving research. *The Quarterly Journal of Experimental Psychology, Section A*, 46(1), 83–118. <https://doi.org/10.1080/14640749308401068>
- Dienes, Z., & Berry, D. (1997). Implicit learning: Below the subjective threshold. *Psychonomic Bulletin & Review*, 4(1), 3–23. <https://doi.org/10.3758/BF03210769>
- Dörner, D. (1981). Über die Schwierigkeiten menschlichen Umgangs mit Komplexität. *Psychologische Rundschau*, 32, 163–179.

- Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Hans Huber.
- Dörner, D., Schaub, H., Stäudel, T., & Strohschneider, S. (1988). Ein System zur Handlungsregulation oder – Die Interaktion von Emotion, Kognition und Motivation. *Sprache & Kognition*, 7, 217–232. https://doi.org/10.1007/978-3-642-75304-6_9
- Frensch, P. A., Haider, H., Rüniger, D., Neugebauer, U., Voigt, S., & Werg, D. (2003). The route from implicit learning to awareness of what has been learned. In L. Jiménez (Hrsg.), *Attention and implicit learning* (S. 335–366). John Benjamins Publishing Company. <http://dx.doi.org/10.1075/aicr.48.17fre>
- Funke, J. (1986). *Komplexes Problemlösen – Bestandsaufnahme und Perspektiven*. Springer. <https://doi.org/10.1007/978-3-642-70994-4>
- Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung*. Springer. [urlhttps://doi.org/10.1007/978-3-642-77346-4](https://doi.org/10.1007/978-3-642-77346-4)
- Haider, H. (1992). Implizites Wissen und Lernen. Ein Artefakt? *Zeitschrift für Experimentelle und Angewandte Psychologie*, 39, 68–100.
- Haider, H., Eichler, A., & Lange, T. (2011). An old problem: How can we distinguish between conscious and unconscious knowledge acquired in an implicit learning task? *Consciousness and Cognition*, 20(3), 658–672. <https://doi.org/10.1016/j.concog.2010.10.021>
- Hayes, N. A., & Broadbent, D. E. (1988). Two modes of learning for interactive tasks. *Cognition*, 28, 249–276. [urlhttps://doi.org/10.1016/0010-0277\(88\)90015-7](https://doi.org/10.1016/0010-0277(88)90015-7)
- Kluwe, R. H., Misiak, C., & Haider, H. (1990). Learning by doing in the control of a complex system. In H. Mandl, E. de Corte, N. Bennett, & H. F. Friedrich (Hrsg.), *Learning and instruction. European research in an international context: Bd. 2.1. Social and cognitive aspects of learning and instruction* (S. 197–218). Pergamon Press.
- Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, 119(3), 264–275. <https://doi.org/10.1037/0096-3445.119.3.264>
- Putz-Osterloh, W. (1985). Selbstreflexionen, Testintelligenz und interindividuelle Unterschiede bei der Bewältigung komplexer Probleme. *Sprache & Kognition*, 4, 203–216.

- Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, 6(6), 855–863. [https://doi.org/10.1016/S0022-5371\(67\)80149-X](https://doi.org/10.1016/S0022-5371(67)80149-X)
- Reber, A. S., & Lewis, S. (1977). Implicit learning: An analysis of the form and structure of a body of tacit knowledge. *Cognition*, 5(4), 333–361. [https://doi.org/10.1016/0010-0277\(77\)90020-8](https://doi.org/10.1016/0010-0277(77)90020-8)
- Rose, M., Haider, H., & Büchel, C. (2010). The emergence of explicit memory during learning. *Cerebral Cortex*, 20(12), 2787–2797. <https://doi.org/10.1093/cercor/bhq025>
- Shanks, D. R., & St John, M. F. (1994). Characteristics of dissociable human learning systems. *Behavioral and Brain Sciences*, 17(3), 367–447. <https://doi.org/10.1017/S0140525X00035032>
- Tubau, E., López-Moliner, J., & Hommel, B. (2007). Modes of executive control in sequence learning: From stimulus-based to plan-based control. *Journal of Experimental Psychology: General*, 136(1), 43–63. [urlhttps://doi.org/10.1037/0096-3445.136.1.43](https://doi.org/10.1037/0096-3445.136.1.43)
- Vester, F. (1974). *Neue Dimensionen des Denkens*. S. Fischer.
- Wessel, J., Haider, H., & Rose, M. (2012). The transition from implicit to explicit representations in incidental learning situations: More evidence from high-frequency EEG coupling. *Experimental Brain Research*, 217(1), 153–162. <https://doi.org/10.1007/s00221-011-2982-7>
- Wundt, W. (1907). Über Ausfrageexperimente und über die Methoden zur Psychologie des Denkens. *Psychologische Studien*, 3, 301–360.

Implicit Learning Across Isoproblems in the Tower of London Puzzle is Impaired in Patients with Parkinson's Disease

AMORY H. DANEK

Ruprecht-Karls-Universität Heidelberg

Abstract. Shallice's Tower of London is popular for assessing problem-solving. One feature of the Tower of London puzzle is the existence of a special class of problems: isomorphic problems, as yet little studied, require identical moves, but the colors of the balls are permuted. Thus, difficulty is the same even if the problems look different. Prior work suggests that implicit learning occurs when a series of isomorphic problems is solved ("iso-effect"). In a computerized version of the task, problem-solving performance by 9 patients with idiopathic Parkinson's disease was compared to healthy controls with regard to several measures of performance, including planning and execution time, moves, efficiency and errors. In general, patients needed more time and more moves to solve the problems, and their solutions were less efficient. Although both groups dedicated the same amount of their time to planning, controls found better solutions. Furthermore, controls showed learning across a set of five isomorphic problems (less time, fewer moves, increasingly efficient solutions), an effect demonstrably different from general learning. In contrast, patients did not improve their isomorphic problem performance. We thus confirmed an iso-effect in healthy participants and suggest that it may be related to implicit memory which is thought to be affected by Parkinson's disease.

Problem-solving research in Germany has largely moved away from static problems to dynamic problems. Joachim Funke was the key player who defined

and explored the new research field of complex problem-solving. As early as 1983 (now 40 years ago!), he investigated the relationship between intelligence and performance in the complex problem-solving scenario “Tailorshop” for which he had re-written the code in the programming language BASIC to run on a Commodore computer (Funke, 1983). His proposal to describe complex problems either as linear structural equations or finite state automata has been very influential (Funke, 2001) and triggered a wealth of empirical studies. Today, the field of complex problem-solving is very prolific, full of interesting findings, novel approaches and discussions (Dörner & Funke, 2017).

Yet while dynamic, complex problems are closer to “real-world” problem solving, some static problems have become very popular testing tools. Classical examples are tower-transformation tasks such as the Tower of Hanoi (ToH) and the Tower of London (ToL), which are still widely used today. Their characteristics as well-defined problems with clear start and goal states and operators make them a valuable tool for neuropsychological testing and diagnosis (e.g., Carlin et al., 2000; Goel & Grafman, 1995; for a review, see Sullivan et al., 2009). They are also used to distinguish different problem-solving strategies such as hill-climbing or means-end-analysis (Dunbar, 1998). This area of research has seen important methodological advancements, too. First, improved formulations of their mathematical structure and the corresponding “state graphs”, i.e., graphical representations of the entire problem space (Hinz et al., 2005). Second, the development of computerized versions of tower problems which enable the collection of a large number of variables of interest (e.g., solution times, number of moves, errors, efficiency). In particular, some software tools now include the possibility to map each move onto the state graph which allows a fine-grained analysis of problem solvers’ trajectories through the problem space (Hinz et al., 2009; Kaller et al., 2011).

The present study implements the ToL which was originally designed by Tim Shallice (1982). To solve the ToL, problem solvers have to rearrange three differently colored balls that are placed on three rods from an initial configuration to match a certain goal configuration. There are two restrictions: only one ball at a

time can be moved and only three, two, or one ball(s) can be placed on each of the three rods of descending lengths, respectively (see Fig. 1).

An interesting feature of the ToL is the existence of isomorphic problems (“isoproblems” for short), first recognized by Berg and Byrd (2002) and already discussed by Unterrainer et al. (2005). Isoproblems require identical moves, as they have the same start and goal position, but the colors of the balls are permuted. Past research indicates that problem solvers are not aware of their similarity (Johns, 2005). The present work focuses on this special sort of problem by exploring a peculiar learning effect that was found in a previous study (Faber et al., 2009). The so-called “iso-effect” describes an improvement in performance from one isoproblem to the next. In the Faber et al. study, 81 healthy volunteers were presented with a series of 22 problems in fixed order, with six isoproblems interspersed. Participants’ performance improved significantly with respect to all variables investigated (solution time, initiation time, completion time, moves, and efficiency),

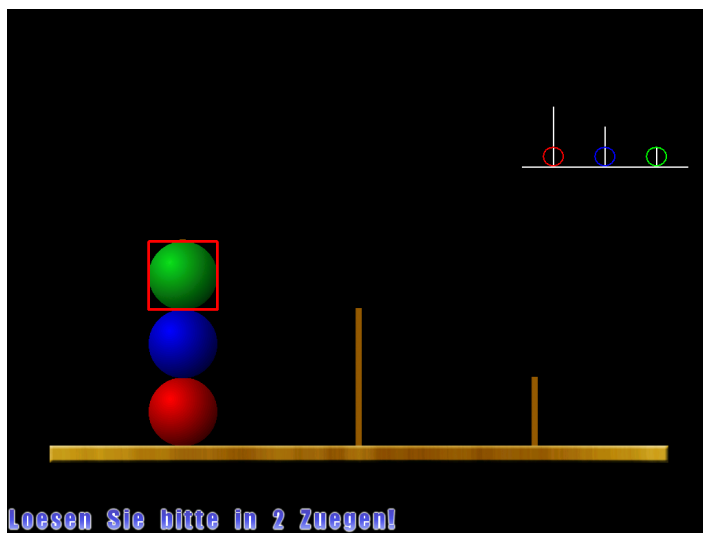


Figure 1. Screenshot of the computerized Tower of London version used in this study. The goal state is depicted in the upper right corner.

but only one third of the sample was aware of the repetition of these problems. From a comparison with other, non-isomorphic problems, it was concluded that the effect seems to be specific to isoproblems and distinct from general learning.

Based on this prior study, it was hypothesized that some kind of implicit learning takes place while solving isoproblems, since problem solvers are not aware of solving a series of problems that require identical moves. Hence, an exploratory study with a small sample size was conducted, testing a patient group which would likely have problems with implicit memory, namely, patients with idiopathic Parkinson's disease. If this patient group would not show any improvement across isoproblems, this would be a first indication that the iso-effect might actually be based on implicit learning. The present study could then serve as a starting point for further, higher-powered studies.

Parkinson's disease (PD) is a progressive, idiopathic neurological disorder related to the loss of dopaminergic cells in the substantia nigra. In addition to motor impairments, cognitive deficits are observed during disease progression, among them deficits in problem-solving and tasks related to executive functions. PD affects the basal ganglia and has previously been connected to impairments of implicit memory (Ruitenberg et al., 2015). The basal ganglia receive projections from the substantia nigra with dopamine as neurotransmitter, so a loss of dopaminergic cells in the substantia nigra may lead to impairment of implicit memory. The ToL has often been used in patient populations. Specifically for PD patients, performance deficits have been reported (e.g., Lange et al., 1992; Owen et al., 1995). Culbertson et al. (2004) compared a large group of PD patients to healthy controls and found longer solution times, more moves, and more rule violations in the Tower of London-Drexel (ToL^{DX}, an improved variant of the original ToL with slight modifications).

The main goal of this exploratory study is to examine the impact of isoproblems on problem-solving performance and to replicate the previously reported learning effect ("iso-effect", Faber et al., 2009). Since the iso-effect is probably based on implicit learning, it should be absent or reduced in participants with impaired implicit learning such as patients with PD. It is thus hypothesized that the iso-effect

will recur in a healthy control group, but not in PD patients. The hypothesis that there is a lack of awareness for isoproblems will also be addressed. Finally, the feasibility of a newly developed ToL tablet version for PD patients will be assessed. At issue is whether patients in advanced stages of the disease will be able to master the cognitive demands of the task and whether their motor impairments still allow them to control the game by using a touchscreen and a special pen.

Method

Participants

A sample of 20 participants was tested, ten Parkinson patients and ten healthy controls. The patients were recruited by screening the outpatient files of the Department of Neurology, and the patients' spouses served as controls. One participant from each group had to be excluded from the analysis: the control person because of too-advanced age (20 years older than the next-oldest participant) and the patient because she differed extremely from the other patients with regard to her severely impaired motor skills. Including her would have compromised the homogeneity of the patient group. This results in a final sample size of $n = 9$ for each test group. In the patient group, age ranged from 51 to 74 years ($M = 66.1$, $SD = 7.42$). In the control group, the range was from 40 to 71 ($M = 59.0$, $SD = 10.67$). A t-test confirmed that groups were matched in terms of age, $t(16) = 1.64$, $p = .12$, and gender, $t(16) = 1.46$, $p = .17$.

Characterization of Patient Sample

The diagnosis of idiopathic Parkinson's disease was established by neurologists who also rated the patients for severity of symptoms using the Hoehn and Yahr scale. The following additional information was collected: duration of disease, prescribed drugs, last intake and dose. Patients were also asked if they had an implanted brain stimulator.

Problem-Solving Task

The Tower of London is a transformation task. Participants are required to rearrange a set of three balls placed on three rods from their initial position into another specified goal configuration. Rods differ in height and can therefore hold only one, two or three balls, respectively. The balls are colored blue, red and green. A computerized version of the game was used for testing which had been developed by Hinz et al. (2009) after the model of the original wooden game by Shallice (1982). The physical appearance, the instructions given, and the selection of problems was strictly oriented on Shallice's original version. A screenshot is shown in Figure 1. This ToL version runs on an easy-to-handle tablet PC that can also serve as a clinical bedside test. The balls are moved from one rod to the other by tapping on the touchscreen with a pen. Participants select a ball which is then highlighted with a red frame. In order to move the ball to another rod, they simply tap at the chosen rod. The ball is then moved automatically from one rod to the other. For each problem, the minimum number of moves required to solve it is shown below the start configuration on the screen and participants are asked to solve the task in the indicated number of moves.

Selection of Problems

The optimal solution path ranges from one to a maximum of eight moves in any possible ToL problem. The ToL problem space contains 210 problems that are spatially unique (Berg & Byrd, 2002), as they differ from all other problems in at least one move in their optimal pathway. Every one of them has five “twins”

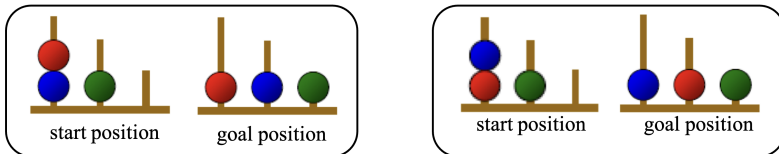


Figure 2. *Two Tower of London isoproblems*

which are called “isoproblems” (Berg & Byrd, 2002). Isoproblems require identical moves as they have the same start and goal position, but the colors of the balls are permuted (see Fig. 2). Those sets of isoproblems have the same level of difficulty (assuming there is no color preference whatsoever) but look different to the observer.

The selected problems corresponded to Shallice’s original range of problems of two to five moves, see Appendix for a detailed list. Eleven problems were implemented. The isoproblems were taken from a set with five moves. To keep the total number of problems small, only five isoproblems were taken from the set (although one iso-set always consists of six problems). The chosen isoproblem was explicitly mentioned by Shallice (1982, p. 204) as being one of the harder problems.

Goal hierarchy: ToL problems can also be categorized according to their start and goal configurations (see Fig. 3). Depending on the configuration used, the sequence of goal moves may be completely ambiguous (flat), partially ambiguous (partial tower) or unambiguous (tower). That is the reason why problems with the same minimum number of moves but different structural properties require different planning abilities and thus have different levels of difficulty (Kaller et al., 2011). As flat configurations seem to be the most difficult ones (Kaller et al., 2004), they were not used in the present study in order to keep difficulty at a level appropriate for patients.

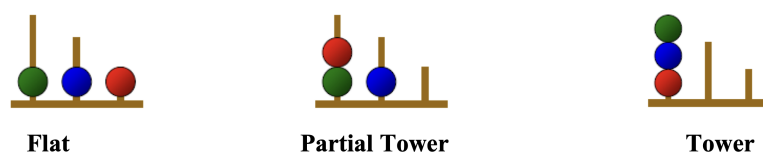


Figure 3. *The three possible start and goal configurations for the Tower of London*

Isoproblems Compared with a Non-isomorphic Control Problem

In addition to the set of five isoproblems, one more five-move problem was selected (see Appendix, Problem #8, non-isomorphic to the other five-move problems). The idea was to compare the isoproblems to another 5-move problem with similar difficulty. The challenge was to find a comparable problem, because the difficulty of a problem is hard to define. Therefore, a problem from the previous study (Faber et al., 2009) was chosen, based on the performance of a sample of 40 healthy controls; see Table 1 for a comparison between the isoproblem and the control problem.

Table 1. *Comparison of ToL data between two problems (one now used as isoproblem, the other as control problem) from the Faber et al. study (2009)*

Performance measure	Isoproblem	Control problem
Moves	5.85	5.88
Efficiency Value	0.88	0.92
Solution Time	64.8	61.7

Note. t-tests for repeated measures showed that the two problems did not differ significantly in any performance measure. This data was obtained from 40 healthy controls, aged 46 to 70 years.

Procedure

Participants were tested individually. Instructions were given in written form and further explained by the experimenter. Speed and accuracy were equally stressed. A practice trial of four problems was performed to get used to pen and touchscreen. Participants completed eleven problems in one session of approximately 45 minutes. Directly after the last problem, to check for their awareness of isoproblems, participants were asked: “Have you noticed anything concerning the last two problems?” Awareness of isoproblems was assumed if the response was “They followed the same principle” or similar statements. If participants had not noticed anything special, a second question was asked: “Have you maybe noticed that

the last two problems were the same, only with different colors?”. In this case, a positive answer was coded as “awareness after cue”.

Tower of London Performance Measures

ToL problem-solving can be assessed on various dimensions of performance. Berg and Byrd (2002) divide all possible measures into four groups: a) success and accuracy; b) efficiency; c) speed; d) strategy measures and rule breaks. They suggest using a complete set of measures from these categories to obtain a comprehensive estimate of individual performance. Following those recommendations, seven performance measures were chosen. The computerized ToL version automatically recorded those parameters.

- 1. *Solution Time*: Amount of time (in seconds) needed to solve a problem. It consists of two phases: Initiation (for thinking about the problem) and completion (to execute the moves), see Figure 4.
- 2. *Initiation Time*: Time from presentation of a problem to the first move (time needed for planning). As recommended by Berg and Byrd (2002), the Initiation Time period ends with the second tap of the first move (when the problem solver selects the rod where the ball should be placed); see Figure 4. This seems useful because even after having selected a ball to be moved,

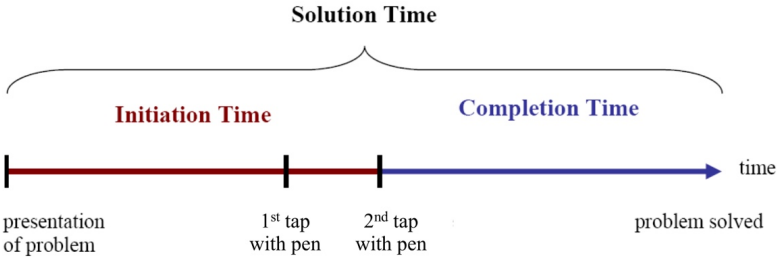


Figure 4. *Solution Time for one problem consists of Initiation Time plus Completion Time*

problem solvers may continue planning until they eventually choose a rod to place that ball on.

3. *Completion Time*: Time from first to last move (i.e., when the solution is completed).
4. *Efficiency Value*: There can be more or less efficient solutions to a problem. Efficiency was calculated by dividing the minimum number of moves by the actual number of moves needed. The Efficiency Value therefore ranges between 0 and 1, with a value of 1 indicating optimal performance.
5. *Moves*: Number of moves needed to solve a problem.
6. *Errors*: Every rule violation (e.g., attempts to put a ball on a rod that is already full) was counted as error.
7. *Awareness of Isoproblems*: Problem solvers' responses to the question about awareness of the last two isoproblems were coded as follows: no awareness, awareness after cue, awareness without cue.

Parameters 1, 2 and 3 measure speed, 4 is a measure of efficiency, and Parameters 5 and 6 provide information about the strategy used.

Results

Awareness of Isoproblems

Most participants (66.7 % of the sample) did not realize that they had solved isoproblems. Only 22.2 % of all participants recognized instantly that the last two problems could be solved with the same sequence of ball movements and could tell that they differed only by a transformation of ball color. Another 11.1 % realized it after they had received an additional cue, as detailed in the procedure. PD patients did not differ from controls in their awareness of isoproblems.

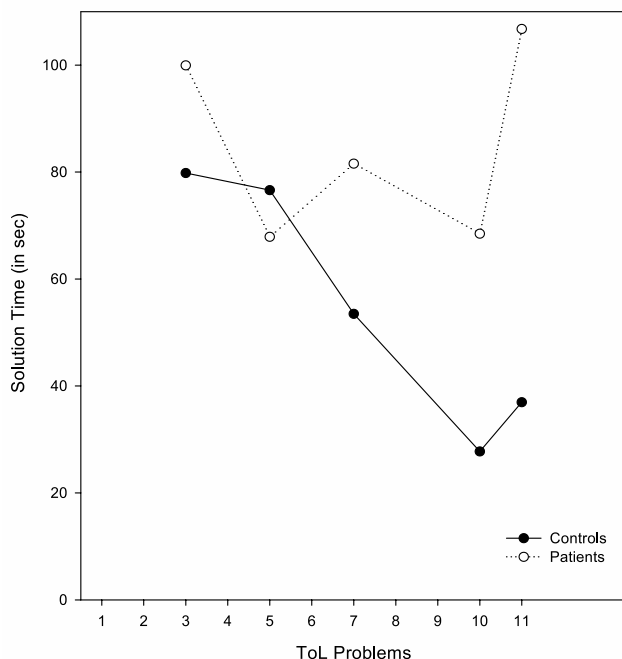


Figure 5. *Mean Solution Time for each of the five isoproblems (controls in black, patients in white)*

Comparison Patients Versus Controls

All eleven problems were solved by all participants. As expected, patients' problem-solving performance was lower compared to controls. A two-way t-test for independent samples showed significant differences between the two test groups with respect to the following variables: Solution Time, $t(16) = 2.80, p = 0.02$; Completion Time, $t(16) = 2.87, p = 0.02$; Moves, $t(16) = 2.34, p = 0.04$; and Efficiency Value, $t(16) = 2.76, p = 0.02$. The difference in committing errors was not significant with $p = 0.056$. Patients had longer overall solution times ($M = 78.6$ secs, $SD = 37.1$) than controls ($M = 42.7$ seconds, $SD = 10.2$) and needed more time (Completion Time: $M = 62.8$ secs, $SD = 35.9$) to execute the movements

than controls ($M = 27.5$ seconds, $SD = 8.5$), but there was no difference between the groups with regard to the variable Initiation Time where both groups had a mean value of 15 seconds. Patients required more moves ($M = 7.2$, $SD = 2.3$) than controls ($M = 5.4$, $SD = 0.6$) and had a lower efficiency ($M = 0.74$, $SD = 0.1$) in solving problems than controls ($M = 0.88$, $SD = 0.1$).

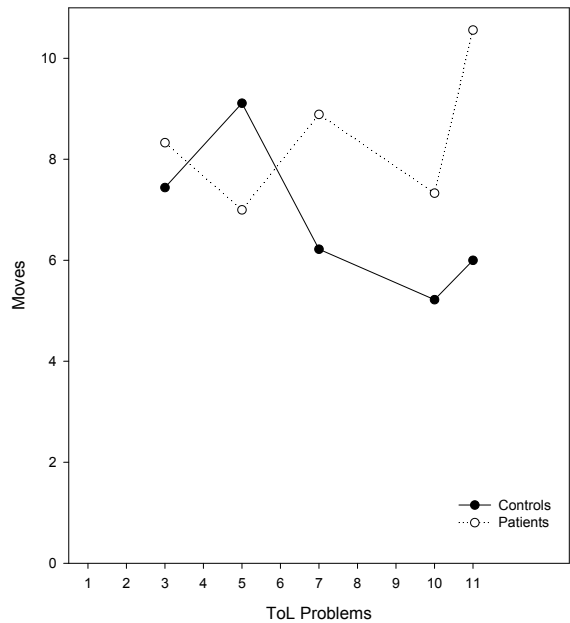


Figure 6. *Mean number of moves for each isoproblem (controls in black, patients in white)*

Iso-Effect

Separately for each test group, performance across the five isoproblems was compared via ANOVA for repeated measures. There were no significant differences between the five isoproblems for any performance variable in the patient group,

supporting the hypothesis that no iso-effect would occur for PD patients. In contrast, controls showed improved performance across the five isoproblems, with significant effects for the following variables: Solution Time, $F(4, 32) = 2.78, p = 0.04, \eta_p^2 = 0.26$; Moves $F(4, 32) = 3.23, p = 0.03, \eta_p^2 = 0.29$; and Efficiency Value, $F(4, 32) = 3.97, p = 0.01, \eta_p^2 = 0.33$. Controls became faster from isoproblem to isoproblem. In Figure 5, the learning curve of controls can be compared to the W-shaped performance of patients who did not improve their performance.

A similar iso-effect, shown in Figure 6, could be observed for the number of moves needed to reach a solution. Likewise, participants from the control group were able to find increasingly efficient solutions across isoproblems; see Figure 7.

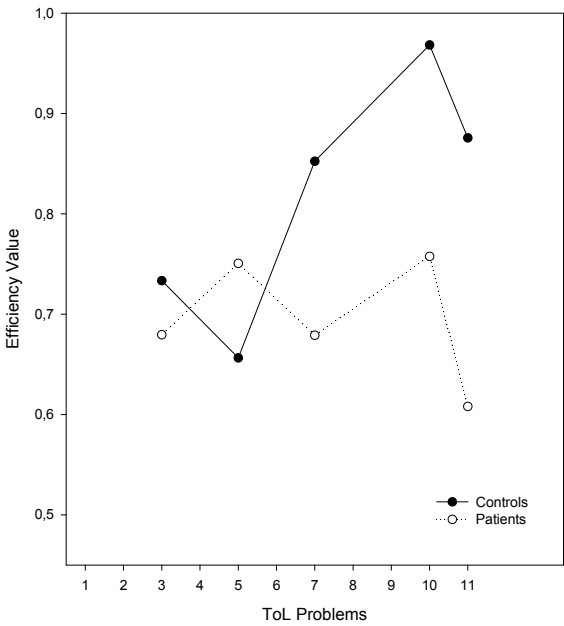


Figure 7. Mean efficiency for each isoproblem (controls in black, patients in white)

Isoproblems Compared with a Non-isomorphic Control Problem

As described in the section on problem selection, one five-move control problem was presented in addition to the set of five-move isoproblems. The idea was to compare performance for isoproblems with another five-move problem of similar difficulty. In Figure 8, mean Solution Times for the isoproblems are plotted together with the non-isomorphic control problem. If the observed decrease in Solution Times across isoproblems was caused by a general learning effect, the presentation of another, non-isomorphic problem of the same difficulty should not interrupt the learning curve. But the control problem (marked by an arrow) does not fit onto the

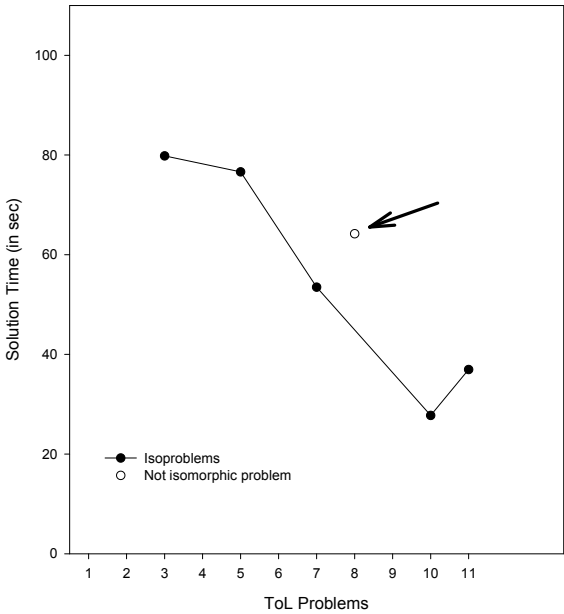


Figure 8. *Solution Time learning curve for isoproblems. The non-isomorphic problem (marked by an arrow) does not fit onto the curve.*

curve—instead there is a drop in performance, reflected in longer Solution Times. This can be taken as an indication that the iso-effect is specific to isoproblems and not due to general learning. The same pattern is observed for the variables Moves and Efficiency (Fig. 9 and 10). The data presented here is based only on control participants because, for patients, no iso-effect and, hence, no learning curve was found.

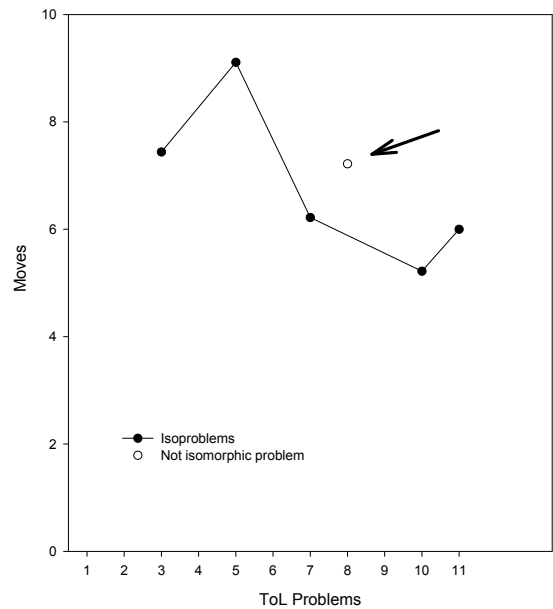


Figure 9. Move learning curve for isoproblems. The non-isomorphic problem (marked by an arrow) does not fit onto the curve.

Discussion

In the present exploratory study, the previously found iso-effect (Faber et al., 2009) occurred again for three main ToL performance measures. Control participants

learned from isoproblem to isoproblem with regard to solution time, moves and efficiency. They needed less time, fewer moves and their solutions became increasingly efficient. This is in contrast to patients, who did not improve their performance in any variable.

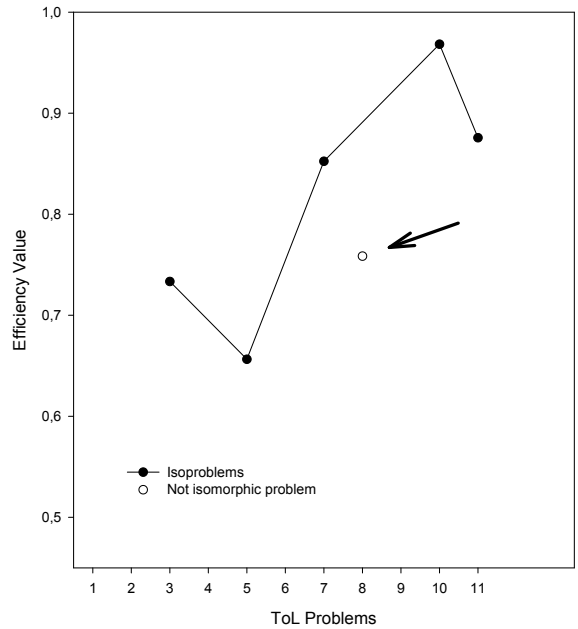


Figure 10. *Efficiency learning curve for isoproblems. The non-isomorphic problem (marked by an arrow) does not fit onto the curve.*

In accordance with prior work (Faber et al., 2009; Johns, 2005), only one third of the sample was aware of the repetition of isoproblems, supporting the hypothesis of a lack of awareness for isoproblems. Thus, it seems that problem solvers learned without realizing it. It has been suggested (Berg & Byrd, 2002; Kaller et al., 2011) that isoproblems could be very useful for repeated assessment. Implementing isoproblems in subsequent testing sessions offers the possibility of comparing

participants' performance on functionally identical problems. In the present study, most participants were not aware of the similarity of these problems, although the repetition took place within minutes. It can therefore be assumed that, in different testing sessions with more time in between, there would most likely be no awareness of solving the same problems again and again. This demonstrates the great value of isoproblems for neuropsychological testing.

While only based on a very small sample, the present results seem to support our main hypothesis and prediction: the iso-effect recurred in the healthy control group but not in the patient group. Why did PD patients not show the learning effect? Why could they not benefit from the repeated presentation of the same problem, in contrast to controls? Is this deficit specific to PD patients, or would other patients show the same pattern? Further research using higher-powered studies and including other patient groups is needed to answer these questions. For now, the present findings are regarded as a first indication that isoproblems do actually measure implicit learning and that the reason why PD patients did not show the iso-effect lies in their impairment of implicit memory.

It is important to distinguish the iso-effect from general learning caused by familiarity with the task and by solving a series of eleven problems. It is argued here that the iso-effect is distinct from general learning in two ways: It is based on implicit learning and occurs after very few repetitions (the second isoproblem was solved significantly better than the first one). Evidence for the first claim is given by the fact that most participants showed no awareness of solving the same problem and by the finding that PD patients (i.e., patients with impaired implicit memory) did not show the effect. The learning effect observed by Johns (2005) occurred while a series of 48 problems was solved, but participants in the present study improved while solving only five isoproblems.

The comparison with a non-isomorphic control problem further supports the idea that the iso-effect is specific to isoproblems and not due to a general learning effect as proposed by Johns (2005). If the improvement from isoproblem to isoproblem was caused by general learning, the presentation of another, non-isomorphic problem of the same difficulty should not interrupt the learning curve,

since problem solvers should be able to benefit from prior exposure to comparable problems and, therefore, improve their performance further. Instead of such continuous improvement, a drop in performance for the non-isomorphic problem was observed.

As expected, the control group performed better overall than the patient group. However, this comparison yielded some findings that deserve further discussion. The differences in solution time and completion time between PD patients and controls should not be interpreted to the effect that patients needed more time to think about the problem. As there was no significant difference in initiation time, it seems reasonable that patients only needed more time to execute the movements—and that this caused the difference in overall solution time, which consists of initiation and completion time.

Interestingly, the same pattern of results was obtained in the previous study (Faber et al., 2009) when comparing younger with older participants. Older participants (like the patients) needed more time to execute the movements, but did not differ from younger participants with regard to initiation time. In that study, it was concluded that motor skills rather than cognitive skills seemed to distinguish the young from the old in the computerized version of the ToL task. This conclusion must be reconsidered, because in contrast to the older control group who was as good as the younger one in finding efficient solutions, patients differed substantially from the other group in the quality of their solutions. For patients, it is not just motor skills that distinguish them from controls, but they also show cognitive deficits resulting in less optimal solutions.

The present results suggest that PD patients are comparable to healthy controls with respect to the amount of planning done. However, the same amount of planning does not result in the same quality of solutions—an unexpected finding which requires discussion. On average, the efficiency of controls was much higher than the efficiency of patients, and approached the optimal value of 1, indicating that controls solved more problems in the minimum number of moves. Although both groups dedicated the same amount of time to planning, controls were able to find more efficient, better solutions. Does the length of the planning phase have no

impact on the quality of solutions? Did patients decide for the wrong solution paths during their planning phase? Or did patients not plan at all before executing the first move, but merely used this time for general orientation (identifying the goal configuration, thinking about the rules...)? Judging from the informal observation of patients' behaviour during the task, the last possibility seems to be the most plausible one. Obviously, other experimental procedures, such as thinking-aloud, are needed to back up these observations. In any case, this interesting finding needs to be further investigated in future studies.

Another aspect to consider is the true nature of the deficits shown by PD patients. Which underlying functions are impaired? Is it really a cognitive deficit? It could also be argued that deficits in visuospatial perception cause the poor performance, since visuospatial impairments have been reported in PD patients (Boller et al., 1984; Kerai et al., 2012; Ransmayr et al., 1987). For further research, it is thus recommended to implement an additional test to assess patients' visuospatial abilities.

Obviously, due to the small sample size, the present study must be interpreted with caution. It was intended to test whether the iso-effect would be found again and whether PD patients would be a good target population for investigating this learning effect. As it stands, the present results are encouraging and will hopefully serve as starting point for further studies into the iso-effect and the possible underlying mechanism of implicit memory.

A final question was whether the newly developed ToL tablet version (where the game is controlled via touchscreen and special pen) would be feasible for PD patients. In the present sample, all patients were able to master the motor demands of the task. However, this finding might be limited to PD patients with a deep brain stimulator (seven out of nine patients had one implanted). Considering the fact that the brain stimulator considerably decreases the tremor, it remains unclear if patients without this treatment would be able to complete the task. However, the two patients from our sample without a brain stimulator succeeded as well.

In accordance with previous studies (Culbertson et al., 2004), the effectiveness of the ToL in differentiating between patients with Parkinson's disease and healthy

controls could be demonstrated. PD patients took more time to solve the problems, they required more moves and their solutions were less efficient compared to controls. Hence, the ToL puzzle on a tablet PC could also contribute to the identification of PD patients.

Acknowledgements. I would like to thank Professors Andreas M. Hinz, Adrian Danek, and Kai Bötzel, Ludwig-Maximilians-Universität, München, for guidance about the mathematical properties of the ToL, for making the computerized version of the ToL available for testing, for valuable advice and inspiring discussions and for providing access to PD patients. Furthermore, I am grateful to my colleagues Dr. Fatma Sürer and Nicole Kühnpast for helpful discussions about Tower tasks. I also thank all participants, especially the patients, who volunteered to take part in the study.

Author Note. The present text is based on the Bachelor's thesis "Isoprobabilities in the Tower of London puzzle: A means to assess implicit memory?" written by Amory H. Danek in 2006.

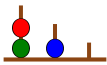
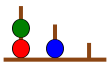
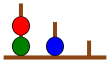
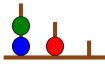
References

- Berg, W. K., & Byrd, D. (2002). The Tower of London spatial problem-solving task: Enhancing clinical and research implementation. *Journal of Clinical and Experimental Neuropsychology*, 24(5), 586–604. <https://doi.org/10.1076/jcen.24.5.586.1006>
- Boller, F., Passafiume, D., Keefe, N. C., Rogers, K., Morrow, L., & Kim, Y. (1984). Visuospatial impairment in Parkinson's disease: Role of perceptual and motor factors. *Archives of Neurology*, 41, 485–490. <https://doi.org/10.1001/archneur.1984.04050170031011>
- Carlin, D., Bonerba, J., Phipps, M., Alexander, G., Shapiro, M., & Grafman, J. (2000). Planning impairments in frontal lobe dementia and frontal lobe lesion patients. *Neuropsychologia*, 38(5), 655–665. [https://doi.org/10.1016/s0028-3932\(99\)00102-5](https://doi.org/10.1016/s0028-3932(99)00102-5)
- Culbertson, W. C., Moberg, P. J., Duda, J. E., Stern, M. B., & Weintraub, D. (2004). Assessing the executive function deficits of patients with Parkinson's disease:

- Utility of the Tower of London-Drexel. *Assessment*, 11(1), 27–39. <https://doi.org/10.1177/1073191103258590>
- Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8 (1153). <https://doi.org/10.3389/fpsyg.2017.01153>
- Dunbar, K. (1998). Problem solving. In W. Bechtel & G. Graham (Eds.), *A companion to Cognitive Science* (pp. 289–298). Blackwell.
- Faber, A. H., Kühnpast, N., Sürer, F., Hinz, A. M., & Danek, A. (2009). The Iso-Effect: Is there specific learning of Tower of London iso-problems? *Thinking & Reasoning*, 15(3), 237–249. <https://doi.org/10.1080/13546780902924308>
- Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? *Diagnostica*, 29, 283–302.
- Funke, J. (2001). Dynamic systems as tools for analysing human judgement. *Thinking & Reasoning*, 7(1), 69–89. <https://doi.org/10.1080/13546780042000046>
- Goel, V., & Grafman, J. (1995). Are the frontal lobes implicated in “planning” functions? Interpreting data from the tower of Hanoi. *Neuropsychologia*, 33(5), 623–642. [https://doi.org/10.1016/0028-3932\(95\)90866-p](https://doi.org/10.1016/0028-3932(95)90866-p)
- Hinz, A. M., Klavžar, S., Milutinović, U., Parisse, D., & Petr, C. (2005). Metric properties of the Tower of Hanoi graphs and Stern’s diatomic sequence. *European Journal of Combinatorics*, 26(5), 693–708. <https://doi.org/10.1016/j.ejc.2004.04.009>
- Hinz, A. M., Kostov, A., Kneißl, F., Sürer, F., & Danek, A. (2009). A mathematical model and a computer tool for the Tower of Hanoi and Tower of London puzzles. *Information Sciences*, 179(17), 2934–2947. <https://doi.org/10.1016/j.ins.2009.04.010>
- Johns, N. B. (2005). *The role of Tower of London iso-problems on the problem solving abilities of adults* [Honor’s Thesis]. University of Florida, Gainesville.
- Kaller, C. P., Rahm, B., Köstering, L., & Unterrainer, J. M. (2011). Reviewing the impact of problem structure on planning: A software tool for analyzing tower tasks. *Behavioural Brain Research*, 216(1), 1–8. <https://doi.org/10.1016/j.bbr.2010.07.029>
- Kaller, C. P., Unterrainer, J. M., Rahm, B., & Halsband, U. (2004). The impact of problem structure on planning: Insights from the Tower of London task. *Cognitive Brain Research*, 20(3), 462–472. <https://doi.org/10.1016/j.cogbrainres.2004.04.002>

- Kerai, J. H., Bracewell, R. M., Hindle, J. V., & Leek, E. C. (2012). Visuospatial transformation impairments in Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology*, 34(10), 1053–1064. <https://doi.org/10.1080/13803395.2012.716396>
- Lange, K. W., Robbins, T. W., Marsden, C. D., James, M., Owen, A. M., & Paul, G. M. (1992). L-Dopa withdrawal in Parkinson's disease selectively impairs cognitive performance in tests sensitive to frontal lobe dysfunction. *Psychopharmacology*, 107(2–3), 394–404. <https://doi.org/10.1007/BF02245167>
- Owen, A. M., Sahakian, B. J., Hodges, J. R., Summers, B. A., Polkey, C. E., & Robbins, T. W. (1995). Dopamine-dependent frontostriatal planning deficits in early Parkinson's disease. *Neuropsychology*, 9, 126–140. <https://doi.org/10.1037/0894-4105.9.1.126>
- Ransmayr, G., Schmidhuber-Eiler, B., Karamat, E., Engler-Plöerer, S., Poewe, W., & Leidlmaier, K. (1987). Visuoception and visuospatial and visuorotational performance in Parkinson's disease. *Journal of Neurology*, 235(2), 99–101. <https://doi.org/10.1007/BF00718018>
- Ruitenbergh, M. F. L., Duthoo, W., Santens, P., Notebaert, W., & Abrahamse, E. L. (2015). Sequential movement skill in Parkinson's disease: A state-of-the-art. *Cortex*, 65, 102–112. <https://doi.org/10.1016/j.cortex.2015.01.005>
- Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, 298, 199–209. <https://doi.org/10.1098/rstb.1982.0082>
- Sullivan, J. R., Riccio, C. A., & Castillo, C. L. (2009). Concurrent validity of the tower tasks as measures of executive function in adults: A meta-analysis. *Applied Neuropsychology*, 16(1), 62–75. <https://doi.org/10.1080/09084280802644243>
- Unterrainer, J. M., Rahm, B., Halsband, U., & Kaller, C. P. (2005). What is in a name: Comparing the Tower of London with one of its variants. *Cognitive Brain Research*, 23(2–3), 418–428. <https://doi.org/10.1016/j.cogbrainres.2004.11.013>

Appendix. ToL problems used

Start position	Goal position	Min. moves	Order	Comment
RG/B/0	0/GB/R	2	Nr. 1	Shallice
RG/B/0	BG/R/0	3	Nr. 2	Shallice
		5	Nr. 3	1 st isoproblem, Shallice
RG/B/0	GBR/0/0			
RG/B/0	G/BR/0	3	Nr. 4	Shallice
		5	Nr. 5	2 nd isoproblem, Shallice
GR/B/0	GBR/0/0			
RG/B/0	B/GR/0	4	Nr. 6	Shallice
RB/G/0	BGR/0/0	5	Nr. 7	3 rd isoproblem, Shallice
		5	Nr. 8	Shallice
RG/B/0	GB/R/0			
RG/B/0	BR/0/G	4	Nr. 9	Shallice
BG/R/0	GRB/0/0	5	Nr. 10	4 th isoproblem, Shallice
BR/G/0	RGB/0/0	5	Nr. 11	5 th isoproblem, Shallice

Note. Capital letters refer to the color of the ball. For example, RG/B/0 refers to the ToL state where the red ball is lying on the green ball on the left rod, the blue ball is positioned on the middle rod and the right rod is empty (compare problem Nr. 3).

Metacognition and Climate Change: Why Dealing with Politicized Science Requires Self-Insight

HELEN FISCHER¹ AND DOROTHEE AMELUNG²

¹Leibniz-Institut für Wissensmedien Tübingen; ²Ruprecht-Karls-Universität Heidelberg

Abstract. Climate change is a prototypical example of a complex problem. Here, we argue that metacognition, our insight into the reliability and fallibility of our own knowledge and thought, is a critical aspect in the area of climate change for two reasons: Metacognition is negatively affected by the noisy information environment of climate change; yet it is at the same time required for its solution. By showcasing the importance of metacognition for climate change, the present chapter thereby provides theoretical and empirical arguments for a link between individual-level psychology and one of the biggest challenges facing societies today.

An American saying states that “For every complex problem, there’s a solution that is simple, neat, and wrong.” In this chapter, we argue and present empirical evidence for the claim that, in order to successfully address the complex problem of climate change, humans need appropriate insight into the reliability and fallibility of their own understanding of the world. In other words, we argue for the claim that, to successfully mitigate climate change, human metacognition is an essential skill.

Climate change is a prototypical example of a complex problem. Joachim Funke writes that complex problems are characterized by five critical aspects: *Complexity* of the involved system, which typically comprises a high number of variables; *interconnectivity* of the variables that interact with each other, producing complex causality and temporal *dynamics*; *intransparency* concerning the involved variables and their interactions; and *polytely*, the fact that solving complex problems often requires meeting not just one, but several—and potentially conflicting—goals (Funke, 2012). What makes climate change complex? The climate system itself is complex in that it is composed of a plethora of variables which are *connected* and interact in typically nonlinear ways (Masson-Delmotte et al., 2021). The global and temporal scale, the multitude of effects on the natural, economic and social systems—all of these interacting variables make it difficult to unambiguously predict mid- to long-term consequences. Adding to this, our behavior constantly interacts with the physical problem of climate change, which shapes collective (in)action, and, in turn, feeds back into the climate system. Moreover, any technological or political response strategies such as climate engineering, carbon capture and storage, or mitigation options add further potential for feedback loops, interactions, and, consequently, ambiguities to the mix, as even mere debate about potential deployment of climate engineering technologies or the adoption of any other response strategy may affect current decisions and behavior (see Amelung & Funke, 2015; Burns, 2011; Dana, 2021).

Climate Change as a Complex Problem

Via a document leak, it has become public just how much governments with vested interests in the fossil fuel industry, such as Saudi Arabia, and other stakeholders with agendas directly in conflict with recommended climate action, such as the meat industry, watered down scientifically derived key recommendations in the 2023 Summary for Policymakers of the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC). These stakeholders view climate change through the filter of how it affects their personal agendas rather than a

collective welfare or a sustainability filter—and their combined action has the potential to create a push-back effect on more ambitious climate goals: The leaked document revealed that more than 36,000 submissions were made by governments, corporations, and other stakeholders with the aim of influencing the climate discourse via text change requests for the 2023 IPCC report, the main communication channel of the international climate science community (Rowlatt & Gerken, 2021). While this procedure is an expression of the democratic principles guiding the IPCC's actions, it also becomes clear that interpretations of climate change differ between stakeholders, indicating how economic and political worldviews filter even highly consensual scientific climate change evidence.

How is it possible that perspectives on the problem of climate change can vary so much? This is because climate change is not merely a problem waiting for a solution but, rather, a global phenomenon with implications for almost every area of planetary and human life. As such, it is not only physically observable and measurable; it also is a topic that is highly contested and subject to political agendas. We could also say it is a complex problem that is perceived through our individual mental filters and interpretations (see Amelung & Funke, 2013).

The complexity of climate change has many implications for how our mental models are formed: as predictions of complex effects are based on scientific uncertainty (Latif, 2011) and, sometimes, even contrasting pieces of evidence or interpretations (Sarewitz, 2004), perceptions of these effects may vary based on how uncertainties and contrasts are communicated (Amelung et al., 2016; Pidgeon & Fischhoff, 2011; Schmid-Petri & Arlt, 2016; Weiss, 2003) as well as understood (Broomell & Kane, 2017; Visschers, 2018); how much various sources of communication are trusted (Arbuckle et al., 2015); and what values, beliefs, or social filters existed prior to receiving the communication (Hornsey et al., 2016). Common cognitive biases (Fischer et al., 2019; Zaval & Cornwell, 2016) as well as misinformation campaigns (Lewandowsky et al., 2012) may add to the confusion.

If, for all of these reasons, our perceptions and interpretations of the meaning of climate change are as multifaceted as the physical phenomenon thereof—is it any wonder that we struggle to find consensus on how to deal with it?

Metacognition and Climate Change

Here, we argue that metacognition—our insight into the accuracy of our understanding of, and reasoning about, the problem at hand—is both affected by the noisy and complex information environment of climate change while, at the same time, being required for its solution. Three cognitive reasons suggest such a twofold involvement of metacognition in the domain of climate change.

First, in the domain of politicized science in general and climate change in particular, citizens are not only confronted with an unmanageably high amount of information but also with a mix of accurate information, misinformation, and outright disinformation, that is, misinformation that is spread with the intention to deceive. This noisy information may reduce the accuracy with which citizens can assess their own understanding of the domain: If citizens habitually sample low-reliability evidence, this necessarily reduces metacognitive sensitivity, the extent to which confidence differentiates between correct and incorrect pieces of knowledge (Fischer & van den Broek, 2021; Rollwage & Fleming, 2021). Indeed, research suggests that metacognitive insight into one's distinction of true and false statements about the domain is reduced for climate change compared to non-politicized science, even when controlling for differences in object-level understanding of the domain (Fischer et al., 2019). In this study, we presented two national samples of German citizens with true and false statements about climate change (e.g., "Climate change is mainly caused by a natural variation in sunbeam and volcanic eruption") as well as about non-politicized science (e.g., "Antibiotics kill viruses as well as bacteria"). Results revealed that citizens lacked metacognitive sensitivity for climate change, although they had high awareness for the accuracy and inaccuracy of their knowledge about non-politicized science. These results are in line with a mechanism whereby widespread misinformation about climate change renders citizens "metacognitively confused" about truth and falsehood.

Second, metacognitive evaluation of our cognitions can lead to enhanced humility about their validity: while object-level reflections take our cognitions for

granted, metacognition enables us to take a step back and critically reflect on the appropriateness of our cognitions. Lacking insight into the fallibility of our cognitions, in contrast, manifests itself in unjustified (biased) confidence, and can foster the psychological tendency to interpret novel evidence selectively such that it confirms prior beliefs. Metacognitive insight into the validity of our cognition may, therefore, be necessary to be appropriately attentive to belief-disconfirming evidence about climate science and to use it for belief-updating. This is particularly so since climate change is a highly politicized domain where strong inclinations exist to process novel evidence such that it confirms a desired conclusion, such as reaffirming one's political or economic worldviews (for reviews, see Hornsey, 2020, 2021). Indeed, empirical results support this theoretical reasoning (Fischer et al., 2022; Said et al., 2021): Metacognitive insight into the accuracy of novel evidence about as well as into one's own understanding of climate change predicted belief updating in response to noisy information that could be interpreted to confirm a desired conclusion, such that individuals with more accurate metacognitive insight were less prone to such motivated belief-updating. In one study, we presented a national sample of German citizens with noisy evidence about climate change and tracked their prior beliefs (the beliefs they had before seeing the evidence) and posterior beliefs (the ones they had after seeing the evidence). In addition, we comprehensively measured their knowledge of climate change and their political and economic attitudes. Results showed that, at one particular level of climate change knowledge, citizens with a more accurate metacognitive insight into this knowledge were less likely to update their beliefs in the direction of misleading evidence. In other words, citizens with more accurate metacognitive awareness of their knowledge—and knowledge gaps—were less susceptible to climate-skeptical belief updating in response to noisy evidence.

And third, people use the confidence they have in their object-level beliefs to guide subsequent judgment and behavior (Balsdon et al., 2020; Desender et al., 2018; Fischer et al., 2023). Hence, high metacognitive sensitivity—where confidence matches the accuracy of the underlying belief—should relate to accurate judgment and behavior in relation to climate change with almost analytical ne-

cessity: When beliefs are weighted by appropriate confidence, low-confidence beliefs that are likely to be inaccurate (such as believing that evidence for anthropogenic climate change is a hoax) are down-weighted and less likely to inform subsequent behavior (such as delaying mitigative policy), while high-confidence beliefs that are likely to be accurate (such as that there is overwhelming evidence for anthropogenic climate change, and its risks) are up-weighted and more likely to inform subsequent behavior (such as supporting mitigative policy). Indeed, metacognitive confidence was found to predict climate change beliefs, above and beyond object-level climate change knowledge (Fischer & Said, 2020). Similarly, we found in another national sample of German citizens that the sensitivity with which confidence matched the accuracy of underlying beliefs about COVID-19 predicted compliance with public health measures and willingness to get vaccinated during the pandemic; and this held at one particular level of accuracy of the COVID-19 beliefs. This result hence suggests that the ability to reflect upon and evaluate our own beliefs can help us make decisions that are informed by the best available evidence, rather than only the accuracy of the beliefs themselves.

It is for all of these reasons that we propose that an insight into the accuracy and fallibility of our own understanding of the world—that is, metacognitive ability—is required for humans to successfully mitigate the complex problem of climate change.

Conclusion

Climate change is a prototypical example of a complex problem. By showcasing how metacognition is both affected by the noisy and complex information environment of climate change while, at the same time, being required for its solution, the present chapter highlights the link between individual-level psychology and one of the biggest challenges facing societies today. This chapter hence provides theoretical and empirical arguments for the claim that to *avoid* a solution that is simple, neat, and wrong, human metacognition is an essential skill.

References

- Amelung, D., Fischer, H., Kruse, L., & Sauerborn, R. (2016). Defogging climate change communication: How cognitive research can promote effective climate communication. *Frontiers in Psychology*, 7, 1340. <https://doi.org/10.3389/fpsyg.2016.01340>
- Amelung, D., & Funke, J. (2013). Dealing with the uncertainties of climate engineering: Warnings from a psychological complex problem solving perspective. *Technology in Society*, 35(1), 32–40. <https://doi.org/10.1016/j.techsoc.2013.03.001>
- Amelung, D., & Funke, J. (2015). Laypeople's risky decisions in the climate change context: Climate engineering as a risk-defusing strategy? *Human and Ecological Risk Assessment: An International Journal*, 21(2), 533–559. <https://doi.org/10.1080/10807039.2014.932203>
- Arbuckle Jr, J. G., Morton, L. W., & Hobbs, J. (2015). Understanding farmer perspectives on climate change adaptation and mitigation: The roles of trust in sources of climate information, climate change beliefs, and perceived risk. *Environment and Behavior*, 47(2), 205–234. <https://doi.org/10.1177/0013916513503832>
- Balsdon, T., Wyart, V., & Mamassian, P. (2020). Confidence controls perceptual evidence accumulation. *Nature Communications*, 11(1), 1–11. <https://doi.org/10.1038/s41467-020-15561-w>
- Broomell, S. B., & Kane, P. B. (2017). Public perception and communication of scientific uncertainty. *Journal of Experimental Psychology: General*, 146(2), 286–304. <https://doi.org/10.1037/xge0000260>
- Burns, W. C. (2011). Climate geoengineering: Solar radiation management and its implications for intergenerational equity. *Stanford Journal of Law, Science & Policy*, 4, 39–55. <https://ssrn.com/abstract=1837833>
- Dana, D. A. (2021). Geoengineering and the question of weakened resolve. In Burns, W., Dana, D., & Nicholson, S. J. (Eds.), *Climate geoengineering: Science, law and governance* (pp. 165–181). Springer International Publishing. https://doi.org/10.1007/978-3-030-72372-9_7
- Desender, K., Boldt, A., & Yeung, N. (2018). Subjective confidence predicts information seeking in decision making. *Psychological Science*, 29(5), 761–778. <https://doi.org/10.1177/0956797617744771>

- Fischer, H., Amelung, D., & Said, N. (2019). The accuracy of German citizens' confidence in their climate change knowledge. *Nature Climate Change*, 9(10), 776–780. <https://doi.org/10.1038/s41558-019-0563-0>
- Fischer, H., Herzog, S., Rebitschek, F., Ketzer, M., & Fleischhut, N. (2022). *Metacognitive and cultural cognition accounts jointly explain believing, and spreading of contested information*. PsyArXiv. <https://psyarxiv.com/2n75x>
- Fischer, H., Huff, M., Anders, G., & Said, N. (2023). Metacognition, public health compliance, and vaccination willingness. *Proceedings of the National Academy of Sciences*, 120(43), e2105425120. <https://doi.org/10.1073/pnas.2105425120>
- Fischer, H., & Said, N. (2020). Importance of domain-specific metacognition for explaining beliefs about politicized science: The case of climate change. *Cognition*, 208, 104545. <https://doi.org/10.1016/j.cognition.2020.104545>
- Fischer, H., & van den Broek, K. (2021). Climate change knowledge, meta-knowledge and beliefs. In A. Franzen & S. Mader (Eds.), *Research Handbook on Environmental Sociology*. Edward Elgar Publishing. <https://doi.org/10.4337/9781800370456.00015>
- Funke, J. (2012). Complex problem solving. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 682–685). Springer. <https://ssrn.com/abstract=2233373>
- Hornsey, M. J. (2020). Why facts are not enough: Understanding and managing the motivated rejection of science. *Current Directions in Psychological Science*, 29(6), 583–591. <https://doi.org/10.1177/0963721420969364>
- Hornsey, M. J. (2021). The role of worldviews in shaping how people appraise climate change. *Current Opinion in Behavioral Sciences*, 42, 36–41. <https://doi.org/10.1016/j.cobeha.2021.02.021>
- Hornsey, M. J., Harris, E. A., Bain, P. G., & Fielding, K. S. (2016). Meta-analyses of the determinants and outcomes of belief in climate change. *Nature Climate Change*, 6(6), 622–626. <https://doi.org/10.1038/nclimate2943>
- Hulme, M. (2009). *Why we disagree about climate change: Understanding controversy, inaction and opportunity*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511841200>
- Latif, M. (2011). Uncertainty in climate change projections. *Journal of Geochemical Exploration*, 110(1), 1–7. <https://doi.org/10.1016/j.gexplo.2010.09.011>
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful

- debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., & Zhou, B. (Eds.) (2021). *Climate change 2021: The physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. <https://doi.org/10.1017/9781009157896>
- Pidgeon, N., & Fischhoff, B. (2011). The role of social and decision sciences in communicating uncertain climate risks. *Nature Climate Change*, 1(1), 35–41. <https://doi.org/10.1038/nclimate1080>
- Rollwage, M., & Fleming, S. M. (2021). Confirmation bias is adaptive when coupled with efficient metacognition. *Philosophical Transactions of the Royal Society B*, 376(1822), 20200131. <https://doi.org/10.1098/rstb.2020.0131>
- Said, N., Fischer, H., & Anders, G. (2021). Contested science: Individuals with higher metacognitive insight into interpretation of evidence are less likely to polarize. *Psychonomic Bulletin & Review*, 1–13. <https://doi.org/10.3758/s13423-021-01993-y>
- Sarewitz, D. (2004). How science makes environmental controversies worse. *Environmental Science & Policy*, 7(5), 385–403. <https://doi.org/10.1016/j.envsci.2004.06.001>
- Schmid-Petri, H., & Arlt, D. (2016). Constructing an illusion of scientific uncertainty? Framing climate change in German and British print media. *Communications*, 41(3), 265–289. <https://doi.org/10.1515/commun-2016-0011>
- Vischers, V. H. (2018). Public perception of uncertainties within climate change science. *Risk Analysis*, 38(1), 43–55. <https://doi.org/10.1111/risa.12818>
- Weiss, C. (2003). Expressing scientific uncertainty. *Law, Probability and Risk*, 2(1), 25–46. <https://doi.org/10.1093/lpr/2.1.25>

Pragmatics—A Rich Domain for the Study of Meta-Reasoning

KLAUS FIEDLER¹ AND SHIRA ELQAYAM²

¹Ruprecht-Karls Universität Heidelberg; ²De Montfort University, UK

Abstract. Pragmatic reasoning tasks reveal that meta-reasoning is not only essential as a quality check of higher-order mental performance but also for the selection of appropriate strategies to solve problems for which more than a single normatively correct solution exists. Our analysis shows that pragmatic rules often dominate the logic of other normative rules, as illustrated in the context of the Wason-Selection Task. Various other paradigms corroborate the importance of meta-reasoning for strategy selection to disambiguate pragmatically uncertain problems that allow for different pragmatic interpretations. The remainder of this chapter elaborates on Peter Wason's warning to refrain from too restrictive, overdetermined theoretical accounts of findings for which simpler explanations are available, framing of decisions and choice, the ratio bias and the related denominator neglect, anomalies in causal-impact judgments, metacognitive myopia in advice taking, and the ultimate sampling dilemma. We believe that meta-reasoning is sorely needed in present and future research on complex problem solving tasks, as depicted in the work of Joachim Funke (2010, 2014).

Several decades ago, one of the authors felt enormous resistance against the argument that language can be causally antecedent to cognition (Semin & Fiedler, 1988), quite in the spirit of linguistic relativity (Fiedler, 2008a). In the meantime, the one-sided dogma that language causally reflects cognition but not

vice versa, gave rise to a more flexible meta-theory, which covers the impact on judgments and decisions of verbal quantifiers (Budescu & Wallsten, 1995; Teigen & Brun, 2000), framing effects (Kahneman & Tversky, 2000), lexical priming (Fiedler et al., 2005), and saying-is-believing effects (Higgins & Rholes, 1978). Yet, in these paradigms, the role of language is almost exclusively confined to priming basic mental functions. The message conveyed in the present article says that linguistic pragmatics can even determine mental activities at the highest level, namely, metacognitive regulation of reasoning and problem solving activities.

Two Major Functions of Meta-Reasoning

Modern approaches (Ackerman & Thompson, 2017; Koriati, 2019; Petty et al., 2007; Sternberg, 2021) have broadened the domain of metacognitive research to include meta-reasoning in addition to meta-memory. The resulting novel perspective of meta-cognitive research is no longer confined to studying such “classical” metacognitive cues as fluency, ease of retrieval, subjective confidence studied in the context of paired-associate learning, or elementary judgments of almanac tasks. Rather, it is now generally acknowledged that the highest level of systematic mental operations—captured by the term reasoning—calls for similar quality checks as elementary learning and memory tasks. In other words, meta-reasoning involves monitoring and control (Nelson & Narens, 1994) of reasoning inferences – conceived as an insurance of thinking, reasoning and problem solving (Funke, 2010, 2014) against slips of attention and superficial, error-prone thought processes.

The present article goes one more decisive step further, focusing on a second major function of meta-reasoning. In addition to the a-posteriori detection and correction of questionable or invalid reasoning results, an equally prominent function of meta-reasoning is strategy selection. Given that more than one solution path is available to tackle a problem, a super-ordinate meta-reasoning perspective is required to select the best-suited strategy. The remainder of this chapter is devoted to this highly prominent function of meta-reasoning.

To tackle the super-ordinate aim of meta-cognitive regulation, it is first essential to understand that more than a single normative solution exists for most reasoning problems, including scientific reasoning. Consider, for illustration, the manner in which people estimate the believability of cognitive inferences or statements. How likely is it that an ingratiating compliment reflects an honest feeling or judgment, rather than a deceptive attempt?

Language-driven strategy selection. On the one hand, believability and veracity judgments could depend strongly on lexical cues and the context of verbal communication. Research on the linguistic category model (Fiedler, 2008b; Semin & Fiedler, 1988) has shown that language users' truth judgments and their verification inferences depend crucially on the linguistic tools used to express social behavior (i.e., the compliment). Verifying a manifest action verb (to compliment, to ingratiate, to flatter, to insult) depends crucially on situation and person information (e.g., whether it involves *strangers* or *intimate friends*, taking place in the *office* or in the *bedroom*). In contrast, an effective way to verify trait adjectives is to assess the semantic consistency or similarity of multiple traits (e.g., whether a person described as *friendly* is also *rewarding*, *kind*, *benign*, *cheerful*, and *not unfair*). Thus, the use of lexical categories can determine the selection of cognitive inference rules for verifying the truth or credibility of communicated statements.

Pragmatic World Knowledge, Beyond Lexical Meanings

The focus of the present chapter is on the role of pragmatic world knowledge in meta-reasoning, conceived as selection of problem solving strategies in line with a language culture's sensitivity for pragmatic context information, which is to a considerable extent wired into the lexicon and into a sophisticated set of situated communication rules.

Social conventions versus normative knowledge. In a meta-analysis of verbal, non-verbal, and para-verbal cues used in over 150 lie-detection studies, Hartwig and Bond (2011) demonstrated that social conventions can dominate normative rules. Applying Zuckerman, DePaulo, and Rosenthal's (1981) lens-model analysis

to lie detection, Hartwig and Bond (2011) assessed each cue's ecological validity (i.e., the cue's actual diagnosticity or statistical relationship to the objective truth criterion) and the coefficient of cue utilization (i.e., the weight given to each cue in participants' subjective truth judgment). While all ecological validities were quite modest (see filled black bars in Fig. 1), reflecting the limited ecological validity of cues to lie detection, the most impressive and most telling result of the meta-analysis was that cue-utilization coefficients (open bars) were regularly higher than ecological validities. Human lie detectors' clichés or collective agreement about the weighting of cues in subjective truth judgments were stronger by magnitudes than the cues' actual diagnostic value. For instance, there was wide agreement among participants that gaze aversion or object fidgeting (e.g., nervous playing around with jacket buttons or pens) are valid indicators of lying, although the actual diagnostic value of these cues was close to zero.

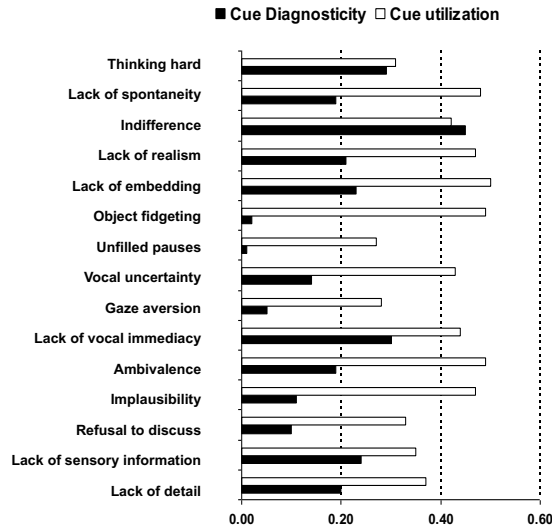


Figure 1. *Ecological validities (cue diagnosticities) and cue-utilization coefficients of prominent cues to lie-detection in a meta-analysis by Hartwig & Bond (2011)*

Meta-Reasoning and Normative Selection

Thus, collective conventions about what seems to discriminate a lie from the truth create social realities, the importance of which should not be underestimated. Regardless of the cues' "true" diagnosticities, an applicant in a job interview or a defendant in a court trial should seek eye contact and avoid object fidgeting, to prevent lie detectors from drawing wrong but highly predictable inferences.

Pragmatically, then, knowledge of socially shared clichés can be more important than the normatively correct cue diagnosticities. Adaptive meta-reasoning entails world knowledge (or meta-knowledge) about social conventions when knowledge about unquestionable normative criteria is not available. Meta-reasoning calls for a realistic insight that it is hardly possible to distinguish lies from truthful communication. And still, in distinct pragmatic contexts, when it comes to practical advice and instructions for job applicants or eyewitnesses one has to recognize that even invalid cues can be of key instrumental importance for trust and persuasion. What determines success of job applicants, defendants, and witnesses is subjective credibility rather than normatively determined "objective" credibility. Thus, the domain of meta-reasoning transcends the domain of normative criteria. By pointing to the social conventional nature of "reality", the meta-reasoning perspective also transcends the confines of cognitive psychology and highlights political surplus meanings of conformity, opposite to Hannah Ahrendt's (1963) reminder of the obligation to be disobedient (Fiedler, 2021).

What we have discussed so far is nothing but a prelude to understanding the crucial role of meta-reasoning in logical reasoning tasks, as inspired by decades of fruitful research in the tradition of the so-called *Wason Selection Task* (WST; Wason, 1968; Evans & Over, 2004, Chapter 5)—one of the most studied tasks in the history of cognitive psychology. The WST was first conceived as a test of the Popperian idea that science progresses by refuting rather than confirming hypotheses. The aim of the present chapter is to demonstrate that the meta-reasoning perspective offers enlightening solutions to even the most prominent paradigm of cognitive psychology, as well as to its most crucial meta-theoretical question, the

question of normativity. Given multiple normative criteria, how is the appropriate norm to be chosen? The same insights will be the focus of later sections dealing with other reasoning paradigms.

In the classical, abstract version of the WST, participants are presented with four cards, with numbers on one side, and letters on the other side. Only one side is exposed. Their task is to test a conditional rule of the form ‘If p , then q ’, for example, ‘If there is an A on one side of a card, then there is a 6 on the other side’. The cards typically show the values for p , $not-p$, q and $not-q$; for example, A, H, 6, 8, respectively. Participants’ task is to turn over all the cards—and only the cards—necessary to determine if the conditional rule is true or false. The normatively correct solution, according to propositional logic, is to turn over the p card and the $not-q$ card—in this case, A and 8, since these are the cards that can refute the rule (A in case there is something other than 6 on the other side, and 8 in case there is an A on the other side). Typically, no more than 10 percent of participants select this correct double option. Most participants *match* (Evans, 1972)—they select the cards mentioned in the rule, in this case A and 6, the p card and the q card, respectively. The most prevalent responses by far are either p alone, or p and q . Participants who do select the p and $not-q$ cards tend to be more cognitively able and motivated (Stanovich, 1999). Matching cards are selected faster than non-matching cards, eliciting higher *Feelings of Rightness* (Thompson et al., 2013).

A flurry of experimental variations followed, with various takes on versions that facilitated performance. For a while, research appeared to favor a *thematic facilitation* hypothesis, the idea that participants perform better when the materials are concrete rather than abstract, but this interpretation was soon undermined by spectacular failures to replicate. It rather turned out that for the conditional rule to be understood, it should be *deontic*—it should depict socially meaningful regulations. Deontic logic is the logic of permissions and obligations, with logical operators such as ‘should’, ‘must’, ‘allowed’ and ‘forbidden’ (McNamara, 2022). The most well-documented deontic version of the selection task is the *drinking age* task, using the rule ‘If a person drinks beer, then this person is over 19 years

of age' (Griggs & Cox, 1982), with cards depicting a person on one side and a beverage on the other side. For the cards Beer, Coke, 22, and 16 (p , $not-p$, q , $not-q$, respectively), about 75% of participants selected the normatively correct p , $not-q$ (Beer, 16) cards. They easily understood that they have to check whether people who drink beer are under 19 and that 16-year old (or younger) people must not drink beer. The same goes for Cosmides's (1989) social contract task ('If a man eats cassava root then he must have a tattoo on his face'), the deontic meaning of which is only understood if participants know pragmatically that cassava is an aphrodisiac and that in this culture only married men have a face tattoo.

The WST offers a good opportunity to examine normativity issues in the context of meta-reasoning. The relevant logic for the deontic task is not the normative rule of propositional reasoning but deontic logic – which is pragmatically more appropriate for this kind of task. As a prominent example, consider Oaksford and Chater's (1994) Bayesian interpretation: when the aim is to select the card providing the optimal data rather than the one refuting the rule, the p and q selections become (given a few extra premises) perfectly normative.

The WST is not unique in casting normative solutions into question. In almost every single research paradigm in reasoning and decision making, more than one normative solution is applicable. This begs the question of *arbitration*: which one normative system should be selected, if at all? One possibility is *descriptivism* (Elqayam, 2012; Elqayam & Evans, 2011): this is the idea that normative considerations do not advance research in the psychology of reasoning. Rather, research is best led by clearly describing the rational act. This does not necessarily mean abandoning the notion of rationality. With a descriptivist approach, whether an act is rational or not is judged by its practical effectiveness, for instance, whether it contributes to the agent achieving her goals (also known as instrumental rationality, or rationality₁, in the terminology proposed by Evans & Over, 1996).

Another possibility is to develop a meta-normative theory tool for strategy selection. This is what Schurz and Hertwig (2019) did. Arguing that intuition alone is a dubious guide to norm selection, they suggested *cognitive success* as an alternative route. Cognitive success is a technical term, defined as the cognitive

component of practical rationality, the rationality of achieving one's goals. Schurz and Hertwig argue that a precondition to practical rationality is the ability to predict how actions lead to consequences. The actual formula to compute cognitive success can be quite complex, involving a trade-off of precision and generalizability. To illustrate, analyses of predictions of the outcomes of football matches demonstrated that a 'take the best' heuristic (i.e., selecting the most superior option in terms of a single most valid cue) outperformed other selection strategies in a pragmatically striking manner: while its precision was suboptimal, its generalizability turned out to be highest. Note that cognitive success as defined by Schurz and Hertwig seems to offer a potentially useful measure for theorists, but it falls short of a *descriptive psychological* model portraying how reasoners select a norm in reality.

A more psychological approach to norm selection can be found in Skovgaard-Olsen et al. (2019), who developed the 'scorekeeper task' to measure norm preference. This task gauges normatively reflected attitudes. It relies on a vignette of a dialogue between fictional protagonists, who are portrayed as participants in a paid experiment. Each protagonist justifies a specific norm as the basis for their responses. The norms were formed according to theories of conditionals (conditionals are sentences of the form *if p, then q*). The fictional protagonists responded either in conformity to a norm based on conditional probability, or based on reason relations. In a norm based on conditional probability, conditionals are evaluated based on the probability of *q* given *p*; for example, for the sentence 'If global warming continues, London will be flooded', we would estimate the probability that London will be flooded given that global warming continues.

In contrast, when the norm is based on reason relations, reasoners test if *p* raises the probability of *q*; in this case, if global warming raises the probability of London being flooded. Participants were instructed to indicate whether payment for participation should be awarded or withheld from the protagonists based on the justifications. What Skovgaard-Olsen et al. show is that responses on a range of tasks are consistent with the specific norm that participants favor in the scorekeeper task. For example, when participants favored the protagonist expressing a reason relation, they tended to disagree that 'If global warming continues, London will

be flooded’ logically follows from ‘Global warming will continue and London will be flooded’. This inference, *p and q; therefore, if p then q* (also known as ‘centering’), is valid for conditional probability, but not for reason relations. The advantage of this paradigm is its sensitivity to individual differences and the potential for accommodating multiple norms. The downside is the lack of psychological explanation of the norm selection *process*; that is, why and how some participants select norm A, whereas others select norm B.

In conclusion, the WST task offers a useful paradigm to study the disambiguation of reasoning norms, both for theorists as well as for reasoners.

In the preceding sections on judgments of credibility and logical correctness, we have already seen how meta-reasoning serves both functions announced at the outset. In addition to the quality check and error correction of one’s own reasoning products, an equally important function of meta-reasoning is strategy selection, that is, selecting of reasonable strategies of problem solving, dependent on the contextualized interpretation of a problem that is essentially ambiguous. A major requisite of the strategy-selection capacity is the emancipated wisdom that more than one reality and more than one best normative rule exist at the same time, depending on the framing of the problem at hand.

Meta-Reasoning Supports Strategy Selection

In this section, we elaborate further on this crucial meta-cognitive ability, with particular reference to the fact that analyzing the same problem at different aggregation levels will often call for different “correct solutions”. Emancipation amounts to understanding and tolerating the co-existence of more than one truth. As in the preceding section, we shall see that a pragmatic perspective on meta-reasoning has the potential to enrich different paradigms of cognitive research on reasoning.

Another lesson from Peter Wason: Consider the following sketch of another piece of seminal research by Peter Wason (1960), in which participants have to discover a rule that underlies a series of stimuli. Let us assume that the sequence ‘1’, ‘2’, ‘4’ has been already confirmed as fitting the rule in question. On each of

the following trials, participants can suggest a possible candidate stimulus, and they receive feedback about whether the suggested stimulus fits the rule or not. A typical response on the next trial would be '8', and the feedback would be positive: 'Yes, 8 indeed fits the rule'. On the next trial, a typical response would be '16', again followed by positive feedback. When, the next suggestion, '32', is again met with confirmatory feedback, the vast majority of all participants are perfectly certain that the rule is 2^N (2 to the power of N).

Indeed, this is a correct solution, but many other, weaker solutions might also be correct. The correct solution might be super-linearly increasing integer numbers, or increasing integer numbers, or simply increasing numbers, real numbers, or maybe numbers, or alphanumerical symbols, or anything. Normally, nobody tested and ruled out the possibility that less restrictive solutions may also be correct. Wason (1960) felt that the preference for too strong overdetermined hypotheses, and the insensitivity for less specific, weaker or broader hypotheses is very common. People fail to understand that simpler and less restrictive hypotheses can also be correct and can provide more parsimonious accounts than more sophisticated hypotheses in a hierarchically ordered set of nested hypotheses.

For an up-to-date example (see Fiedler, Kutzner & Krueger, 2012), the notion of mortality salience relies on the experimental demonstration that if participants are exposed to a funeral, or if they write a short vignette about their own mortality, they exhibit a marked conservative shift in political opinions and in verbal expressions of interests and values (Greenberg et al., 1994; Weise et al., 2008). As more and more experiments provided convergent validation for a conservative shift induced by mortality salience, especially after the nine-eleven trauma in the US, hardly anybody would cast the validity of this interpretation into doubt. But could not the weaker possibility that any existential emotion (e.g., birth of a baby, opposite to death) may also induce a conservative shift, or that other traumatic experiences (unrelated to mortality) may foster conservative values? Would reminding people of their incompleteness (of which mortality is but a special case) have a similar effect, or may simply inducing a Zeigarnik effect (i.e., leaving an experimental task incomplete) be sufficient for a conservative shift (Wicklund & Braun, 1990)?

The mortality example highlights the fact that experimental treatments (like going to a funeral) are never pure treatments of only one independent variable; they always manipulate a whole hierarchy of increasingly general variables, of which the focal variable of interest is only a special case. A mortality treatment is always a special case of incompleteness. Conversely, mortality is a general version of more specific meanings of mortality (related to bodily weakness, survival, religion, specific causes of death). A super-ordinate meta-reasoning function is required to disambiguate the appropriate aggregation level. This meta-reasoning function is not only essential for experimental research on mortality salience but for a wide range of everyday judgments and decisions, to disambiguate health risks, probabilities, life expectancies, or product price levels. For instance, one's personal risk of being the victim of a traffic accident depends crucially on the aggregation level used to estimate our risk (e.g., the entire population or an individual's distinct life setting). No objective norms disburden the meta-reasoning function of the crucial obligation to select an aggregation level.

The pragmatics of framing effects: The notion of framing effects motivated prospect theory (Kahneman & Tversky, 1979, 2000), the most prominent decision theory in behavioral science. For instance, in the so-called Asian-disease problem, participants are to take the role of a health politician who is to choose the better program against a threatening Asian disease:

Program I: 200 hundred people survive for sure

Program II: 1/3 chance that 600 people survive; or else nobody survives

Because of prospect theory's sigmoid value function—saving 600 people is worth less than saving 3 times 200 people—the theory predicts risk-averse choices in a positive survival frame. A majority of all people go for the riskless option; they prefer Program I over Program II. However, when equivalent options are framed in terms of losses (dead people) rather than gains (survival), most participants no longer prefer the save option.

Program III: 400 people die for sure over the risky option

Program IV: 1/3 chance that 0 people die; 2/3 chance that 600 die

Thus, loss framing, unlike gain framing, renders people risk-seeking rather than risk-averse. McKenzie and Nelson (2003) provided an intriguing pragmatic account of framing effects and of the Asian disease problem in particular. Drawing on the analogy to a glass that is both half-full and half-empty, they assumed and corroborated empirically that a common pragmatic interpretation of the phrase 'X% will survive' is that at least X% will survive, but maybe more. So the glass is (at least) X% full. In contrast, the phrase '(100-X% will die)' is commonly interpreted to mean that at least 100 - X% will die, but maybe more. Thus, a meta-reasoning interpretation of the ordinary pragmatic meaning of verbal quantifiers—a typical function of meta-reasoning—offers a simple and straightforward account of framing effects in decision and choice.

Ratio Bias, Denominator Neglect, Proportional Change

The impact of meta-reasoning on higher-order cognitive problem solving is particularly enhanced when no normative accuracy criterion is available, or when normative algorithms are not feasible. Meta-reasoning is then sorely needed to select a boundedly rational heuristic (Simon, 1990). For instance, consider the common phenomenon of ratio bias (Denes-Raj & Epstein, 1994) or the related phenomena of denominator neglect (Reyna & Brainerd, 2008) or sample size bias (Price & Matthews, 2009). When participants are given a choice between two lotteries, Lottery A offering 1/10 chances (out of 10) to win \$10 and Lottery B offering 10/100 chances (out of 100) to win the same amount, the majority of participants typically choose the larger lottery. Although the ratio of the winning numerator divided by the set-size denominator is the same for both lotteries ($1/10 = 10/100$), the larger lottery is more attractive because the larger numerator offers more diverse possibilities to win; the equally larger denominator seems to be neglected.

Many similar findings provide convergent evidence for the preference of the same proportion in a ratio of absolutely larger numbers. Risks appear higher

when framed in ratios of numerically large numbers (Price et al., 2006), mental calculation reflects larger estimated sums when the number of additive components increases (Smith & Price, 2010), or the assessment of changing proportions depends strongly on changes in sample size (Fiedler et al., 2016). Increases in proportional quantities (e.g., in proportions of wins) are easy to detect when absolute sample sizes increase too but not when sample sizes decrease. By analogy, decreases in sample size are easy to detect when sample sizes decrease but not when sample sizes increase. People confuse quality and quantity. They do not understand that having many arguments is not the same as better arguments, or that many Amazon ratings need not be more positive than fewer ratings of other products (Powell et al., 2017).

Ratio biases (or related confusions of proportions and sample size) constitute a major challenge for meta-reasoning. To understand why, imagine the task of assessing in a series of n binary observations the proportion of one focal attribute level, such as winning rather than losing, rainy rather than sunny days, or delayed rather than punctual trains. Assessing the proportion of focal outcomes requires a count of the number of focal outcomes (e.g., k wins) in the numerator, which must be divided by n , the total number of outcomes (sample size) in the denominator. The increment encoded for every elementary observation of a focal outcome is $1/n$. The smaller the sample size, the stronger the impact of elementary outcomes. If sample size is very small (e.g., $n = 2$), the weight given to an elementary outcome is $1/n = 1/2$. If sample size is medium (e.g., $n = 10$), large (e.g., $n = 50$) or very large (e.g., $n = 300$), the incremental weight decreases (to $1/10$, $1/50$, and $1/300$, respectively). So, when experiencing a sample proportion, meta-reasoning can only update a proportional frequency observed in a growing sample when the size of the sample is known beforehand. Because it is unrealistic to assume that people can keep track of the size of all ongoing samples, the meta-reasoning regulation of proportional encoding must resort to a simplifying heuristic.

One radical simplification is to engage in summation rather than averaging, counting only the focal outcomes in the numerator, regardless of the total sample size in the denominator. Summation will of course produce an extremely strong

ratio bias; the same proportion (of, say 80% focal outcomes) will appear much stronger in a large sample of $n = 100$ (yielding 80 focal versus 20 non-focal outcomes) than in a small sample of $n = 10$ (yielding only 8 versus 2 focal outcomes).

More likely than a radical summation strategy is a compromise of summation and averaging, based on the correction of the numerator sum by an estimated n as observed on the previous trial. A testable implication of this algorithm, supported by Fiedler et al. (2016), is that the same sample proportion will subjectively appear smaller (larger) if sample size on the preceding trial was large (small). In any case, judgments of proportional quantities highlight the importance of meta-reasoning heuristics when the present n is unknown beforehand.

Causal-Impact Reasoning

Recent findings on causal-impact judgment offer another intriguing explanation for why proportional quantities are so difficult to administrate. The causal impact of X on Y is a measure of the extent to which a shift in a causal variable (ΔX) induces a shift in an effect variable (ΔY). To measure the impact of a poisonous substance in animal food, one can calculate the ratio ($\Delta Y = \text{number of dead animals}$)/($\Delta X = \text{amount of poison added}$). Causal impact is doubled if the same amount of poison kills twice as many animals. By the same ratio logic, causal impact can be doubled if half the amount of poison kills the same number of animals. Causal impact is thus maximal when a minimal cause (ΔX) produces a maximal effect (ΔY). However, the judgments of causal impact are not maximized when a weak cause induces a strong effect but when a strong cause induces a strong effect (Fiedler et al., 2011; Hansen et al., 2013). Countless submitted articles are rejected because reported effects are too weak, but not when the treatments required to induce the same effect are too strong. If a weak manipulation of ΔX is sufficient to induce a strong effect in ΔY , this is not considered superior to the case of a strong ΔX inducing a strong ΔY , disregarding the underlying proportional logic.

Why should this be so? Why is experienced causal impact maximal when a high ΔX and a high ΔY maximize the covariance but not when a small ΔX and a high ΔY maximize the ratio? A plausible answer is that in many real-life settings, people are jointly engaged in detecting causes and assessing their impact. They cannot experience the impact of a cause that they do not detect in the first place. Thus, if a very subtle remark or a hardly detectable non-verbal gesture is enough to frustrate another person, leading to severe aggression, the impact of the remark or gesture may be high, but it may go unnoticed. The aggression is much more likely to be explained as a causal consequence of a very strong frustration or provocation, the operation of which is highly detectable. Thus, according to this meta-reasoning account, proportional quantities may be so difficult to assess because the need to detect quantities when reasoning about quantities lets covariances dominate ratios.

Meta-Reasoning as a Precondition for Rational Reasoning

Although an extensive discussion of meta-reasoning as a precondition for rational decision and choice is beyond the scope of this article, we will cover at least two aspects of rationality, namely, adaptive advice taking and the “ultimate sampling dilemma”.

Meta-reasoning and advice taking: Almost all informed judgments and decisions in the 21st century—whether in the area of health, law, investment, traveling, consumer choices, or scientific research—depend crucially on peer or expert advice. Hardly any problem can be handled based on our first-hand experience alone. Rather, our decisions and actions depend crucially on second-hand information reflecting our peers’ experience or expert knowledge about health risks, investment chances, legal constraints, ecological costs, or product quality. In other words, virtually all decisions involve advice taking and social communication, making advice taking a truly collective endeavor. Granting that advice givers may not always be honest, cooperative, competent, or in full command of a logical principle (like Bayesian calculus), but may sometimes be dishonest, pursuing their own vested interests, or simply be incompetent and logically incapable, it would be

naïve and irrational to blindly follow each and every advice opinion. To distinguish valid from invalid advice, one of the most prominent meta-reasoning tasks of a rational and mature advice taker is to “separate the wheat from the chaff”, that is, to identify and follow useful advice and to discard and not adopt invalid or deceptive advice. Note that this social-communication skill is nothing but an interpersonal variant of the more general skill to distinguish between trustworthy and error-prone strategies and pragmatic problem framings. For instance, rational advice taking must be sensitive to advice givers’ information sources, their own costs and benefits, to experts’ explanation of their good reasons for advice, to social validation and consistency of multiple advice, and—more generally – to the advice givers’ reasoning, above and beyond the advice proper.

Yet, a growing body of research on meta-cognitive myopia (Fiedler, 2012; Fiedler et al., 2019; Fiedler et al., 2023) reveals that advice takers are often naïve and uncritical, falling prey to obviously invalid advice. For example, even when a medical experts’ health risk estimate is obviously informed by biased or blatantly erroneous study samples, and even when they are explicitly informed and explained that and why an advice is misleading and false, they continue to follow the obviously invalid advice (Fiedler et al. 2019). The impact of metacognitive myopia is reminiscent of perseverance experiments (Ross et al., 1975), in which participants who are explicitly debriefed about a deceptive feedback (e.g., feedback of their utter failure on a test) continue to be influenced by the wrong feedback. Perhaps, an even more compelling analogy can be found in Tversky and Kahneman’s (1974) seminal anchoring experiment, showing that numerical anchors can exert massive influences on participants’ quantitative estimates (e.g., of the number of African nations in the UN), even when anchors are generated randomly by a roulette wheel.

Ultimate sampling dilemma: The “ultimate sampling dilemma” (Fiedler, 2008c) relates metacognitive myopia to the aforementioned WST problems in conditional reasoning. Participants could freely sample any information and as much information as they liked from a knowledge base represented on a computer. They were to play the role of a leading manager of an organization, whose task was

to evaluate prior customers' positive (+) or negative (–) feedback about three providers (P_1 , P_2 , P_3) in two product domains (computers C ; telecommunication T). In the universe of all entries in the knowledge base, the same 2:1 ratio of + to – valence held for all providers and product domains. The distribution of domains was also $2(C) : 1(T)$ and the distribution of providers was $4P_1 : 2P_2 : 1P_3$. Thus, all three distributions were skewed but all contingencies were zero, as the distributions of each variable remained constant across all levels of all other variables. On every trial of the sampling task, participants could not constrain the information search (calling for a random draw from the entire knowledge base), or they could constrain the information search in as many variables they liked. For instance, they could ask for a randomly drawn experience with P_1 (delivering a + or – evaluation in C or T), or for a – experience with P_3 in domain C , etc.

The resulting samples and the participants' evaluations of providers depended dramatically on the hypotheses they were testing. If their task was simply to arrive at comparative evaluations of all three providers, the resulting samples and the final evaluations constantly reflected the preponderance of positive outcomes. However, because the expected number of (predominantly) positive evaluations sampled about P_1 was four times as frequent as for P_3 , and twice as frequent as for P_2 , their final preferences were $P_1 > P_2 > P_3$, accordingly. If they were to test the hypothesis that (the most infrequent) provider P_3 may differ from the other two, they sampled mostly P_3 entries and therefore associated the prevalent + entries with P_3 more than with the other providers. If the task was to figure out the origin of most complaints with or deficits of purchased items, they focused on – outcomes and found that the vast majority of negative outcomes originated in P_1 . Focusing on negative outcomes thus led to a reverse preference order: $P_1 < P_2 < P_3$.

Thus, approaching the same distribution in the data base from different vantage points led to systematically different judgments and preferences, reflecting massive metacognitive myopia effects. Although the actual positivity rates were similarly prevalent for all three providers, participants in the latter condition were blind for the fact that their own biased information search had solicited mainly negative experiences, the majority of which were related to the most frequent provider P_1 .

In other words, metacognitive myopia blinded them for a distinct self-generated sampling bias. Insensitivity for sampling biases constitutes a major deficit in meta-reasoning, affording a massive obstacle on the way to rationality (Fiedler, 2008; Fiedler et al., 2023).

Conclusions

To summarize, our discussion of meta-reasoning in pragmatic problem solving, when different pragmatic framings suggest different normative solutions to the same problem, not only highlighted the crucial role of meta-reasoning in strategy and norm selection. Rather, we have also seen that a meta-reasoning approach is sorely needed to understand reasoning in a world in which the existence of objectively correct norms is an exception rather than a rule. The adaptive function of meta-reasoning in the context of higher-order cognitive problem solving tasks of the type portrayed in Funke's (2010, 2014) work must not be reduced to a quality check based on monitoring and control of actually conducted reasoning processes. In addition to the retrograde quality-check of already conducted problem solving operations, meta-reasoning is at least as essential for strategy selection in the first place and for the disambiguation of a pragmatically uncertain problem situation.

Although our review of pertinent research was far from being exhaustive, our discussion included a variety of prominent findings. We started from the Wason Selection Task and Peter Wason's warning to refrain from too restrictive, over-determined theoretical accounts of findings for which simpler and less restrictive explanations are possible. In the remainder of this chapter, we illustrated the interface of meta-reasoning for high-level cognitive reasoning tasks with reference to various prominent paradigms: the framing of decisions and choice, the ratio bias and the related denominator neglect, anomalies in causal-impact judgments, metacognitive myopia in advice taking, and the ultimate sampling dilemma.

Meta-reasoning and rationality: The evidence reviewed across all these paradigms converges in the conclusion that meta-reasoning is at the heart of a theoretically useful conception of rationality. Although we do not want to prescribe

nominal definitions of central conceptions in behavioral research, we want to point out that defining rationality in terms of pragmatic sensitivity and broadminded strategy selection strikes us as theoretically more fruitful than an orthodox economic definition in terms of payoff maximization. Meta-reasoning is particularly suited to understanding (ir)rationality in everyday behavior. The last decade saw a flourishing of reasoning research into issues such as belief in fake news, anti-science ideation, epistemically suspect beliefs (e.g., Čavojová et al., 2020; Pennycook et al., 2020), and what Stanovich (2009) calls “contaminated mindware”, beliefs that undermine rational pursuit of one’s own interests. Such beliefs correlate with epistemic goals rather than with cognitive ability; for example, Actively Openminded Thinking (Stanovich & Toplak, 2023) is a strong predictor.

Future prospects: We believe that the interface of meta-reasoning and pragmatics bears the potential for future research in many other domains, such as collective rationality, construal-level theory and psychological distance, group decision making, and research on such modern topics as fake news and debunking (Chan et al., 2017; Unkelbach et al., 2019). Performance in all these domains or paradigms depends on the meta-cognitive insight that more than one norm or correct solution may co-exist, and that meta-reasoning is sorely needed to deal with the indeterminacy of such problems.

Last but not least, our meta-reasoning perspective raises some political perspectives about all citizens’ obligation to be sensitive to pragmatic aspects of social contracts, the conventionalized nature of social and political reality, and the validity of information spread on the social media. Although the evidence we have reviewed here does not directly speak to these important political issues, we nevertheless hope that our chapter helps to sensitize people to understand that meta-reasoning constitutes a sorely needed tool for political emancipation.

Author Note. The research documented in the present chapter was supported by research grant Fi 294/30-1 awarded to the first author by the *Deutsche Forschungsgemeinschaft* and by IIAS (Israeli Institute for Advanced Studies) scholarships awarded to both authors.

References

- Ackerman, R., & Thompson, V. A. (2017). Meta-reasoning: Monitoring and control of thinking and reasoning. *Trends in Cognitive Sciences*, 21(8), 607–617. <https://doi.org/10.1016/j.tics.2017.05.004>
- Ahrendt, H. (1963). *Eichmann in Jerusalem: A report on the banality of evil*. Faber & Faber.
- Budescu, D. V., & Wallsten, T. S. (1995). Processing linguistic probabilities: General principles and empirical evidence. *Psychology of Learning and Motivation*, 32, 275–318. [https://doi.org/10.1016/s0079-7421\(08\)60313-8](https://doi.org/10.1016/s0079-7421(08)60313-8)
- Čavojová, V., Šrol, J., & Jurkovič, M. (2020). Why should we try to think like scientists? Scientific reasoning and susceptibility to epistemically suspect beliefs and cognitive biases. *Applied Cognitive Psychology*, 34(1), 85–95. <https://doi.org/10.1002/acp.3595>
- Chan, M. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531–1546. <https://doi.org/10.1177/0956797617714579>
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? *Cognition*, 31, 187–276. [https://doi.org/10.1016/0010-0277\(89\)90023-1](https://doi.org/10.1016/0010-0277(89)90023-1)
- Denes-Raj, V., & Epstein, S. (1994). Conflict between intuitive and rational processing: When people behave against their better judgment. *Journal of Personality and Social Psychology*, 66(5), 819–829. <https://doi.org/10.1037//0022-3514.66.5.819>
- Elqayam, S. (2012). Grounded rationality: Descriptivism in epistemic context. *Synthese*, 189(1), 39–49. <https://doi.org/10.1007/s11229-012-0153-4>

- Elqayam, S., & Evans, J. St. B. T. (2011). Subtracting ‘ought’ from ‘is’: Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, 34(5), 233–248. <https://doi.org/10.1017/s0140525x1100001x>
- Evans, J. St. B. T. (1972). Interpretation and matching bias in a reasoning task. *Quarterly Journal of Experimental Psychology*, 24, 193–199. <https://doi.org/10.1080/00335557243000067>
- Evans, J. St. B. T., & Over, D. E. (2004). *If*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198525134.001.0001>
- Fiedler, K. (2008a). Language: A toolbox for sharing and influencing social reality. *Perspectives on Psychological Science*, 3(1), 38–47. <https://doi.org/10.1111/j.1745-6916.2008.00060.x>
- Fiedler, K. (2008b). The implicit meta-theory that has inspired and restricted LCM research: Why some studies were conducted but others not. *Journal of Language and Social Psychology*, 27(2), 182–196. <https://doi.org/10.1177/0261927x07313656>
- Fiedler, K. (2008c). The ultimate sampling dilemma in experience-based decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 186–203. <https://doi.org/10.1037/0278-7393.34.1.186>
- Fiedler, K. (2012). Meta-cognitive myopia and the dilemmas of inductive-statistical inference. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 57, pp. 1–55). Elsevier. <https://doi.org/10.1016/B978-0-12-394293-7.00001-7>
- Fiedler, K. (2021). A non-populist perspective on populism in psychological science. In J. P. Forgas, W. D. Crano, & K. Fiedler (Eds.), *The psychology of populism* (pp. 174–194). Routledge. <https://doi.org/10.4324/9781003057680>
- Fiedler, K., Freytag, P., & Unkelbach, C. (2011). Great oaks from giant acorns grow: How causal-impact judgments depend on the strength of a cause. *European Journal of Social Psychology*, 41(2), 162–172. <https://doi.org/10.1002/ejsp.750>
- Fiedler, K., Hütter, M., Schott, M., & Kutzner, F. (2019). Metacognitive myopia and the overutilization of misleading advice. *Journal of Behavioral Decision Making*, 32(3), 317–333. <https://doi.org/10.1002/bdm.2109>
- Fiedler, K., Kareev, Y., Avrahami, J., Beier, S., Kutzner, F., & Hütter, M. (2016). Anomalies in the detection of change: When changes in sample size are mistaken for changes in proportions. *Memory & Cognition*, 44(1), 143–161. <https://doi.org/10.3758/s13421-015-0537-z>

- Fiedler, K., Kutzner, F., & Krueger, J. I. (2012). The long way from α -error control to validity proper: Problems with a short-sighted false-positive debate. *Perspectives on Psychological Science*, 7(6), 661–669. <https://doi.org/10.1177/1745691612462587>
- Fiedler, K., Prager, J., & McCaughey, L. (2023). Metacognitive myopia: A major obstacle on the way to rationality. *Current Directions in Psychological Science*, 32(1), 49–56. <https://doi.org/10.1177/09637214221126906>
- Fiedler, K., Schenck, W., Watling, M., & Menges, J. I. (2005). Priming trait inferences through pictures and moving pictures: The impact of open and closed mindsets. *Journal of Personality and Social Psychology*, 88(2), 229–244. <https://doi.org/10.1037/0022-3514.88.2.229>
- Funke, J. (2010). Complex problem solving: A case for complex cognition? *Cognitive Processing*, 11(2), 133–142. <https://doi.org/10.1007/s10339-009-0345-0>
- Funke, J. (2014). Analysis of minimal complex systems and complex problem solving require different forms of causal cognition. *Frontiers in Psychology*, 5, Article 739. <https://doi.org/10.3389/fpsyg.2014.00739>
- Greenberg, J., Pyszczynski, T., Solomon, S., Simon, L., & Breus, M. (1994). Role of consciousness and accessibility of death-related thoughts in mortality salience effects. *Journal of Personality and Social Psychology*, 67(4), 627–637. <https://doi.org/10.1037/0022-3514.67.4.627>
- Griggs, R. A., & Cox, J. R. (1982). The elusive thematic materials effect in the Wason selection task. *British Journal of Psychology*, 73, 407–420. <https://doi.org/10.1111/j.2044-8295.1982.tb01823.x>
- Hansen, J., Rim, S., & Fiedler, K. (2013). Psychological distance and judgments of causal impact. *Journal of Experimental Social Psychology*, 49(6), 1184–1189. <https://doi.org/10.1016/j.jesp.2013.05.012>
- Hartwig, M., & Bond, C. F., Jr. (2011). Why do lie-catchers fail? A lens model meta-analysis of human lie judgments. *Psychological Bulletin*, 137(4), 643–659. <https://psycnet.apa.org/doi/10.1037/a0023589>
- Higgins, E. T., & Rholes, W. S. (1978). “Saying is believing”: Effects of message modification on memory and liking for the person described. *Journal of Experimental Social Psychology*, 14(4), 363–378. [https://doi.org/10.1016/0022-1031\(78\)90032-X](https://doi.org/10.1016/0022-1031(78)90032-X)
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291. https://doi.org/10.1142/9789814417358_0006

- Kahneman, D., & Tversky, A. (Eds.) (2000). *Choices, values, and frames*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511803475.002>
- Koriat, A. (2019). Confidence judgments: The monitoring of object-level and same-level performance. *Metacognition and Learning*, 14(3), 463–478. <https://doi.org/10.1007/s11409-019-09195-7>
- McKenzie, C. R., & Nelson, J. D. (2003). What a speaker's choice of frame reveals: Reference points, frame selection, and framing effects. *Psychonomic Bulletin & Review*, 10(3), 596–602. <https://doi.org/10.1037/e537102012-428>
- McNamara, P. (2022). Deontic logic. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2022 edition). Stanford University. <https://plato.stanford.edu/archives/fall2022/entries/logic-deontic/>
- Nelson, T. O., & Narens, L. (1994). Why investigate meta-cognition? In J. Metcalfe & A. Shimamura (Eds.), *Meta-cognition: Knowing about knowing* (pp. 1–25). MIT Press. <https://doi.org/10.7551/mitpress/4561.003.0003>
- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608–631. <https://doi.org/10.1037//0033-295x.101.4.608>
- Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2020). On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs. *Judgment and Decision Making*, 15(4), 476–498. <https://doi.org/10.1017/s1930297500007439>
- Petty, R. E., Briñol, P., Tormala, Z. L., & Wegener, D. T. (2007). The role of metacognition in social judgment. In A. W. Kruglanski & E. T. Higgins (Eds.), *Social psychology: Handbook of basic principles* (2nd ed., pp. 254–284). Guilford Press.
- Powell, D., Yu, J., DeWolf, M., & Holyoak, K. J. (2017). The love of large numbers: A popularity bias in consumer choice. *Psychological Science*, 28(10), 1432–1442. <https://doi.org/10.1177/0956797617711291>
- Price, P. C., & Matthews, T. V. (2009). From group diffusion to ratio bias: Effects of denominator and numerator salience on intuitive risk and likelihood judgments. *Judgment and Decision Making*, 4(6), 436–446.
- Price, P. C., Smith, A. R., & Lench, H. C. (2006). The effect of target group size on risk judgments and comparative optimism: The more, the riskier. *Journal of Personality*

- and *Social Psychology*, 90(3), 382–398. <https://doi.org/10.1037/0022-3514.90.3.382>
- Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences*, 18(1), 89–107. <https://doi.org/10.1016/j.lindif.2007.03.011>
- Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 32(5), 880–892. <https://psycnet.apa.org/doi/10.1037/0022-3514.32.5.880>
- Schurz, G., & Hertwig, R. (2019). Cognitive success: A consequentialist account of rationality in cognition. *Topics in Cognitive Science*, 11(1), 7–36. <https://doi.org/10.1111/tops.12410>
- Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, 54(4), 558–568. <https://psycnet.apa.org/doi/10.1037/0022-3514.54.4.558>
- Simon, H. (1990). Invariants of human behavior. *Annual Review of Psychology*, 41, 1–19. <https://psycnet.apa.org/doi/10.1146/annurev.ps.41.020190.000245>
- Skovgaard-Olsen, N., Kellen, D., Hahn, U., & Klauer, K. C. (2019). Norm conflicts and conditionals. *Psychological Review*, 126(5), 611–633. <https://doi.org/10.1037/rev0000150>
- Smith, A. R., & Price, P. C. (2010). Sample size bias in the estimation of means. *Psychonomic Bulletin & Review*, 17(4), 499–503. <https://doi.org/10.3758/pbr.17.4.499>
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781410603432>
- Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. Yale University Press. <https://doi.org/10.12987/9780300142532>
- Stanovich, K. E., & Toplak, M. E. (2023). Actively open-minded thinking and its measurement. *Journal of Intelligence*, 11(2), 27. <https://doi.org/10.3390/jintelligence11020027>
- Sternberg, R. J. (2021). Meta-intelligence: Understanding, control, and coordination of intelligence, creativity, and wisdom. In R. M. Holm-Hadulla, J. Funke, & M. Wink (Eds.), *Intelligenz – Theoretische Grundlagen und praktische Anwendungen*.

- Heidelberger Jahrbücher Online* (Vol. 6, pp. 473–488). Heidelberg University Publishing. <https://doi.org/10.17885/heiup.hdjbo.2021.1.24399>
- Teigen, K. H., & Brun, W. (2000). Ambiguous probabilities: When does $p=0.3$ reflect a possibility, and when does it express a doubt? *Journal of Behavioral Decision Making*, 13(3), 345–362. [https://doi.org/10.1002/1099-0771\(200007/09\)13:3%3C345::AID-BDM358%3E3.0.CO;2-U](https://doi.org/10.1002/1099-0771(200007/09)13:3%3C345::AID-BDM358%3E3.0.CO;2-U)
- Thompson, V. A., Evans, J. S., & Campbell, J. I. D. (2013). Matching bias on the selection task: It's fast and feels good. *Thinking & Reasoning*, 19(3–4), 431–452. <https://doi.org/10.1080/13546783.2013.820220>
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157), 1124–1131. <https://doi.org/10.1126/science.185.4157.1124>
- Unkelbach, C., Koch, A., Silva, R. R., & Garcia-Marques, T. (2019). Truth by repetition: Explanations and implications. *Current Directions in Psychological Science*, 28(3), 247–253. <https://doi.org/10.1177/0963721419827854>
- Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *The Quarterly Journal of Experimental Psychology*, 12, 129–140. <https://doi.org/10.1080/17470216008416717>
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, 20, 273–281. <https://doi.org/10.1080/14640746808400161>
- Weise, D. R., Pyszczynski, T., Cox, C. R., Arndt, J., Greenberg, J., Solomon, S., & Kosloff, S. (2008). Interpersonal politics: The role of terror management and attachment processes in shaping political preferences. *Psychological Science*, 19(5), 448–455. <https://doi.org/10.1111/j.1467-9280.2008.02108.x>
- Wicklund, R. A., & Braun, O. L. (1990). Creating consistency among pairs of traits: A bridge from social psychology to trait psychology. *Journal of Experimental Social Psychology*, 26(6), 545–558. [https://doi.org/10.1016/0022-1031\(90\)90055-q](https://doi.org/10.1016/0022-1031(90)90055-q)
- Wüstenberg, S., Greiff, S., & Funke, J. (2012). Complex problem solving—More than reasoning? *Intelligence*, 40(1), 1–14. <https://doi.org/10.1016/j.intell.2011.11.003>
- Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology*, 50(1), 304–315. <https://doi.org/10.1037/a0032982>

Zuckerman, M., DePaulo, B. M., & Rosenthal, R. (1981). Verbal and nonverbal communication of deception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 14, pp. 1–57). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60369-X](https://doi.org/10.1016/S0065-2601(08)60369-X)

Can Self-Explanations and Thinking Aloud Promote Performance at the Tower of Hanoi Problem?

CHRISTINE BLECH

FernUniversität in Hagen

Abstract. An experiment ($N = 89$) with a computerized version of the Tower of Hanoi (ToH) investigated the combined effects of instructed verbalization and self-reported expertise on problem solving performance in a 2 (ToH experts vs. novices) $\times$ 3 (verbalization instruction) between-subjects design. The hypothesis that self-explanations can promote ToH performance as compared to silent thinking or conventional thinking aloud was confirmed for ToH experts, but not for novices. Experts who had to explain and justify each move needed more time, but fewer moves to complete the ToH problems than silent or conventional aloud-thinkers. In contrast, the novices needed more time and more moves when engaging in self-explanations. Based on the Cognitive Load Theory, the conflicting demands of verbalization, schema acquisition, and goal-oriented solution behavior are discussed, explaining the novices' inefficient solutions and their presumed working memory overload. Concerning the experts' performance, the positive impact of mental anticipation is highlighted.

One hundred years before cognitive psychologists in Germany discussed the controversial question of whether and how intelligence affects complex

problem solving (Dörner et al., 1983; Funke, 1983), in 1883, the French mathematician Édouard Lucas (see Hinz et al., 2018) worked out a puzzle that would tease the brain of both test subjects and researchers even two turns of centuries later. The Tower of Hanoi (ToH) was born. To psychologists, the ToH is known as a well-defined, simple, versatile problem solving paradigm with applications in research fields as diverse as child psychology (Resing et al., 2020), psychology of old age (Zinke et al., 2014), clinical psychology (Knapp et al., 2017), neurosciences (Anderson et al., 2005), and, of course, general and experimental cognitive psychology (e.g., Welsh & Huizinga, 2005).

The Tower of Hanoi: Problem or Routine Task?

The ToH does not require elaborate equipment; the rules are plain and clear (see, e.g., Simon, 1975): Given three pegs and a varying number of disks, different in size, a pyramid-like stack of disks has to be transferred from one peg to another. Only one disk, a disk from the very top of a stack, may be moved at a time. A larger disk may never be placed on top of a smaller disk. Besides this original “tower-to-tower” version of the ToH, there are also variants in which the initial state and/or the goal states are in “flat” positions or display partial towers (for examples see, e.g., Anderson et al., 2005, or the Method section below). While the haptic, physical variant of the ToH is suitable for studies with specific samples, e.g., children (Resing et al., 2020), the conceptually equivalent computer-based implementation has some advantages for experiments with healthy adult participants. Sequences of different, single ToH problems can be presented quickly in a standardized manner, correct and wrong moves are recorded automatically along with the solution times (Mataix-Cols & Bartrés-Faz, 2002), and even online testing would be feasible.

Compared to real-life problems, the ToH seems artificial and low in complexity. Following the taxonomy by Dörner (1976), the ToH is a prototypical interpolation problem: the *clarity of the goal state* is high (i.e., no goals need to be defined or negotiated by the problem solver) and the *awareness about available means* to reach the goal is high as well. With sufficient exercise, the problem-defining barrier

between given state and goal state (Frensch & Funke, 1995) disappears. That is why in research literature, the tower of ToH is sometimes classified as a measure of planning (e.g., Knapp et al., 2017) or executive functions (Zinke et al., 2014) rather than as a problem solving paradigm. Anderson et al. (2005) trained participants with various ToH subtasks (each with five disks and an optimal move number of 28) to learn the so-called “sophisticated perceptual strategy” in the sense of Simon (1975). After two days, all participants could solve the tasks virtually flawlessly in a predictable manner. This enabled the authors to map distinct, involved brain areas for phases of planning and representation (e.g., encoding positions of disks, setting subgoals), phases of memory activity (e.g., remembering a plan), and phases of motor execution (e.g., moving disks).

On the other hand, the problem solving perspective on the ToH has shown that, as long as participants approach the ToH “naturally”, i.e., untrained, differential solution strategies (for a classification, see Simon, 1975) can be identified in behavioral data and self-reports, leading to a considerable amount of individual variation in performance (Welsh & Huizinga, 2005). Sometimes, when solutions reach a dead end and moves have to be reversed, solvers may even experience conflicts (Jostmann & Gieselmann, 2014), similar to the goal conflicts in complex problems (Dörner et al., 1983). So, it is important to consider problem solvers’ previous experience with the ToH when assessing their performance.

Thinking Aloud and Self-Explanations

Anderson et al. (2005) mapped problem solving processes in the ToH based on solution times and fMRI analysis. Another approach to access cognitive processes is the method of thinking aloud (van Someren et al., 1994). Thinking aloud means to freely express everything that comes to one’s mind while working on a task or—in this case—a ToH problem. The question of whether verbalization affects thinking routines has been a debate for a long time (see Fox et al., 2011, and the reply by Schooler, 2011). According to Fox et al. (2011), the potential reactivity depends on the abstraction and elaboration of the verbalization. In their meta-analysis

of 94 studies, so-called concurrent verbalization, conceptualized as vocalizing thoughts which would have been inner speech otherwise, showed no effects of reactivity. In contrast, verbalizing specific descriptions or explanations went along with higher performance than free, silent thinking. In an experiment with the water jar problem by Luchins (1942), Blech et al. (2020) found that concurrent thinking aloud neither changed the quality of solution nor the solution time, although one might guess that expressing thoughts would take more time than thinking in silence.

In the study at hand, the particular focus was on the potentially beneficial effects of the more elaborate verbalization techniques like self-explanation methods (Chi et al., 1989; Neuman & Schwarz, 1998). Problem solving took place under three different verbalization conditions: in silence, with concurrent thinking aloud, and with self-explanations, meaning that while working on the ToH, problem solvers had to explain where they were moving the disks and why. The central hypothesis was that self-explanation would enhance the ToH performance as compared to silent problem solving or problem solving with mere concurrent thinking aloud.¹ Additionally, solution times and potential effects of expertise were investigated, i.e., whether problem solvers who were already familiar with the ToH (a) solved the ToH with fewer moves than novices and (b) might benefit either less or more from self-explanation instructions.

Method

Recruitment and Sample

The experiment was planned and conducted in the context of distance teaching. Fifteen students of the FernUniversität in Hagen recruited six participants each from their personal acquaintances and tested them at the students' homes. Rooms and computers were prepared according to standardized guidelines (see Blech, 2022), e.g., distractors like radios and telephones were muted, computers were in

¹ The data collection was part of a student project in 2014. The hypothesis was set up prior to raising the data, but there was no official pre-registration.

full-screen mode, only one participant was tested at a time. By drawing lots (from a pool of six lots, containing codes for three different experimental conditions, each code twice), every student investigator made sure to assign treatment conditions randomly while still aiming for balanced cell sizes.

Altogether, 91 participants were tested, one of them to replace the data of an excluded participant whose experimental condition was not documented properly. During data analysis, another data set was excluded for being an extreme outlier with respect to solution time (average time per task: more than 5 *SD* above the group mean; average time per move: more than 7 *SD* above the group mean). The protocol suggested major interruptions during the test session. Out of the remaining 89 participants, 54% were female, 46% male. The age ranged from 16 to 65 years, the average being 36.62 (*SD* = 13.29) years. Asked about their familiarity with the ToH on a scale from 1 (not familiar at all) to 5 (very familiar), 63% rated themselves as not very familiar (scorings 1 or 2) and will be referred to as novices in the analyses. The remaining 37% (scorings 3, 4, or 5) rated themselves as at least medium-familiar and will be referred to as experts (see Table 1 for further details).

Table 1. *Distribution of expertise across experimental conditions*

Experimental Condition	Novices		Experts		Total <i>n</i>
	<i>n</i>	%	<i>n</i>	%	
Silent control	17	57%	13	43%	30
Thinking aloud	23	74%	8	26%	31
Self-explanation	16	57%	12	43%	28
Total	56	63%	33	37%	89

Note. Novices rated themselves at 1 or 2, and experts rated themselves at 3, 4, or 5 on the ToH familiarity Likert scale. Although the frequencies were not perfectly balanced, the test of stochastic dependency between expertise and experimental condition was not significant, $\chi^2(2) = 2.59$, $p = .27$.

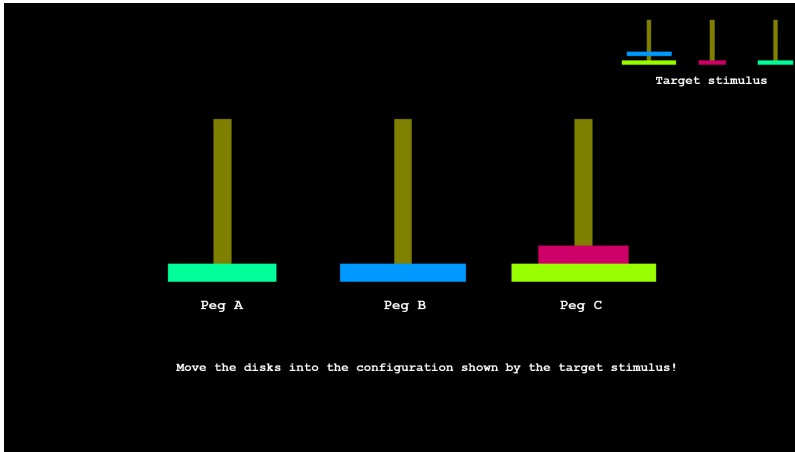


Figure 1. User interface of the ToH software. Translated to English from the original German version. The disks were moved using a computer mouse. The optimal move number is 8.

Material

The experimental software (Hertel, 2014), based on C ++, contained ten computerized ToH problems: Two practice problems with four disks and optimal (i.e., minimal) move numbers of 7 and 8, followed by the eight regular problems with four disks and optimal moves from 8 (problem 1) to 15 (problem 8). The problems were “flat-to-flat” towers presented in order of ascending difficulty, i.e., with an increasing number of optimal moves. An example is given in Figure 1; the full software is available at the Open Science Framework project (Blech, 2022). For all problems, the background color was black, text color was white, and the disks of different sizes varied in color to make discrimination between them easier. Disks were moved by using a computer mouse. The ToH standard instruction ran: “Try to reach the target state in as few moves as possible.” The rules outlined in the introduction were explained prior to the first practice task: (a) move only one disk at a time, and (b) it is not possible to place a larger disk on top of a smaller disk. A sample sketch of a ToH configuration illustrated the position of actual and target position and how to move disks from one peg to another. After the practice task

participants reassured that they had understood the rules. Otherwise, they could ask the experimenter for help before continuing with the main part of the experiment.

Design

Depending on the experimental condition, the ToH standard instruction was complemented by a verbalization instruction manipulated between subjects in a one-factorial design. Participants of the thinking-aloud condition were instructed: "Try to reach the target state in as few moves as possible. Speak out any thoughts that are going through your mind during this time." For participants of the self-explanation condition, the instruction was: "Try to reach the target state in as few moves as possible. Before each move, explain where you are moving a disk and why." Participants of the silent control condition were only given the standard instruction: "Try to reach the target state in as few moves as possible." Conditions were assigned randomly.

Procedure

In the private rooms of the student experimenters, the desks and computers were prepared according to the guidelines mentioned above. The experimenters welcomed the participants by reading off a standard text stating that the study was part of a students' research project in psychology, that the aim was to investigate cognitive processes during problem solving, and that the duration would be approximately 30 minutes. Participants from the think-aloud and self-explanation condition were informed that they had to verbalize, though the verbalizations themselves were not recorded. After signing the informed consent sheet on paper, participants continued with the ToH problems on a computer. All instructions, including the group-specific verbalization instructions, were embedded in the software. Depending on the randomly (per lot) assigned condition, the experimenter started a different version of the software. A glass of water was provided for all participants with and without verbalization instructions. The experimenters stayed in the room with the participant, listening to the verbalizations and making sure

that the instructions were followed. If think-aloud or self-explanation participants turned silent for five seconds or more, the experimenter reminded them to speak out their thoughts on solving the problem. After completing the ToH problems, participants had to answer the questions on demographical variables and the rating scale on ToH familiarity. Finally, they were fully debriefed. They read and were invited to keep a handout explaining, among others things, the three experimental conditions and the hypothesis, together with the experimenters' contact details and the opportunity to ask any questions. The experimenters made sure the data was properly saved on their local computers and transferred the files to the supervising teacher of the project.

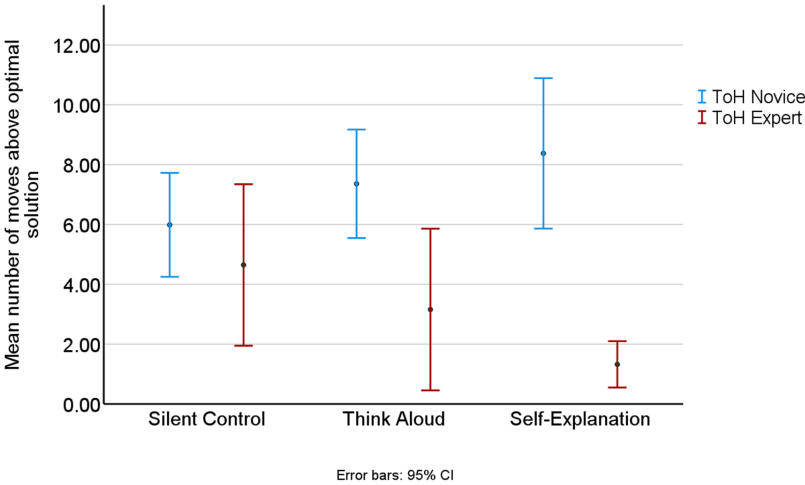


Figure 2. *ToH performance as measured by the average number of moves above the optimum, displayed for silent thinking, thinking aloud, and self-explantation, separated by expertise. Low numbers indicate high performance.*

Measures and Variables

For each ToH problem, the software recorded the number of legal moves and the solution time in seconds exact to three decimal places as raw data. Attempted

illegal moves (mistakes) were also recorded though their execution was blocked: If a participant tried an illegal move, the disk was moved back to its previous position. Practice problems were not included in the analysis. The solution time in seconds was averaged over the eight ToH problems. The average number of legal moves above the optimum served as a measure of performance so that ToH problems requiring more moves had the same weight as simpler problems. According to experimenters' protocols, attempted illegal moves sometimes turned out to be affected by problems with the computer mouse (i.e., not clicking properly, releasing the mouse key too early) and were thus dismissed. SPSS version 28.0.1.1 was used for the statistical analyses. Data, along with an analysis script, is uploaded at the Open Science Framework project (Blech, 2022).

Results

ToH Performance

Figure 2 displays the average numbers of moves above the optimum for the three experimental conditions, separated by expertise. A 3 (verbalization) $\times$ 2 (expertise) between-subjects ANOVA showed a significant main effect of expertise on performance, $F(1, 83) = 23.74$, $p < .001$, $\eta_p^2 = .22$, the experts needing fewer moves than the novices. While the main effect of verbalization was not significant, $F < 1$, ns., the interaction of verbalization $\times$ expertise was significant, $F(2, 83) = 3.90$, $p = .024$, $\eta_p^2 = .09$. Simple effect analyses for the novices did not show a significant main effect for the verbalization condition, $F(2, 83) = 1.61$, $p = .207$, $\eta_p^2 = .04$. In fact, the descriptive results hinted in the direction opposing the hypothesis (see Fig. 2). For the experts, the result pattern matched the expected direction, the number of moves being lowest under the self-explanation instruction and highest for the silent control condition. Yet, the simple effect of verbalization was not significant for the ToH experts either, $F(2, 83) = 2.32$, $p = .105$, $\eta_p^2 = .05$.

Solution times

Paralleling the analyses on the performance measure, a 3 (verbalization) $\times$ 2 (expertise) between-subjects ANOVA was conducted, with solution time as the dependent variable. The average solution time for novices was significantly higher than the solution time for experts, $F(1, 83) = 27.80, p < .001, \eta_p^2 = .25$. The main effect for verbalization became significant as well, $F(2, 83) = 8.09, p < .001, \eta_p^2 = .16$: The longest solution times occurred under the self-explanation condition (see Fig. 3). The interaction of verbalization $\times$ expertise was not significant, $F < 1$, ns. Simple effect analyses for the novices showed a significant main effect for the verbalization condition, $F(2, 83) = 7.67, p < .001, \eta_p^2 = .16$. Unlike with the performance measure, the result pattern for the expert subsample was not opposed to, but, rather, mirrored the result pattern for the novices. Experts, too, had longer average solution times under the self-explanation condition than under the silent control and under the think-aloud condition, although the simple main effect did not become significant, $F(2, 83) = 2.11, p = .127, \eta_p^2 = .05$.

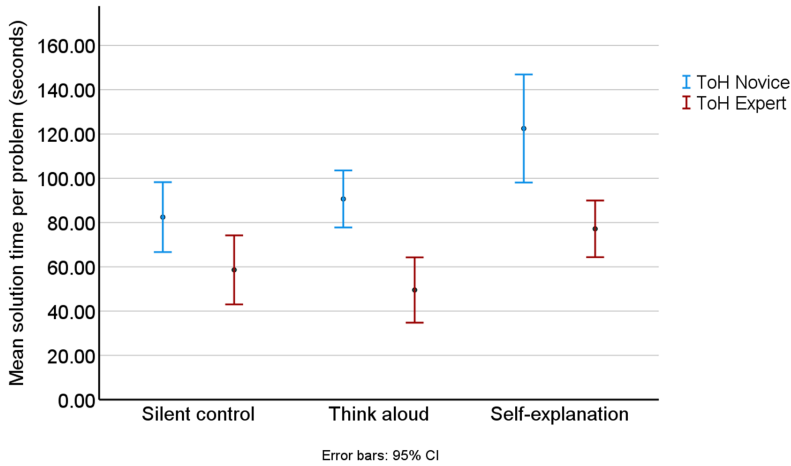


Figure 3. ToH solution times, displayed for silent thinking, thinking aloud, and self-explanation, separated by expertise.

Exploratory Analysis

For explorative purposes, solution times were correlated with the performance measure. On the level of single ToH problems, problems requiring more moves were problems requiring more time, with Pearson correlation coefficients from $r(89) = .60$ for Problem no. 7 and $r(89) = .84$ for Problem no. 4. Similarly, aggregated over all problems per participant, the generally quick problem solvers tended to be the generally accurate problem solvers, the correlation between average number of moves and average solution time being $r(89) = .59, p < .001$.

Discussion

The study reported in this chapter investigated the potential impact of self-explanation and thinking-aloud instructions on the solution performance of a computerized version of the ToH, a simple and frequently applied problem solving paradigm. According to their self-reports of familiarity with the ToH, participants were either classified as experts or novices. A clear, yet not surprising, finding was that on average, experts solved the ToH problems with fewer moves and faster than novices. In the total sample, self-explanations were neither advantageous nor obstructive to the solution performance.

Interestingly, prior experience with the ToH moderated the effect of verbalization instructions: Experts seemed to benefit from self-explanations. They took fewer moves under this condition as compared to silent thinking or concurrent thinking aloud without explanations. Meanwhile, their solution times were not reduced, suggesting that the self-explanation instruction brought the experts to thoroughly anticipate and justify each move before executing it. Similar effects have been reported by Fansher et al. (2022), who compared problem solving of the ToH in a traditional computerized version with a mental imagery condition. The latter condition forced participants to internally simulate the problem solving sequence instead of simply executing the solution step by step, including trial-and-error procedures. As the authors note, in line with the Cognitive Load Theory (Sweller, 1988), mental simulation may consume working memory capacity that would

otherwise have been available for different cognitive processes, such as acquiring new problem solving strategies or schemas. The same effect might occur when forced self-explanations occupy working memory.

Working memory capacity, in turn, is one crucial factor distinguishing experts from novices: By encoding not more, but larger chunks of information (Chase & Simon, 1973) and by using more efficiently organized higher-order problem knowledge (Chi et al., 1982), experts can still have free capacities in problem solving situations that make novices struggle. In the reported experiment, ToH novices performed worst under the self-explanation condition and best under the silent thinking condition. Self-explanations prolonged the solution process but did not improve its outcome. As an interpretation, it is suggested that novices' working memories were overloaded with the combined demand of generating problem solving strategies and putting effort into verbalized self-explanations. As Chi et al. (1989) report, even in traditional thinking-aloud settings, not being prompted to explain solution steps, good problem solvers tend to provide abstracting self-explanations spontaneously, whereas bad problem solvers stick to the mere descriptive level more often. Sweller (1988) puts forward the idea that goal-directed problem solving and the acquisition of knowledge in terms of solution schemes are conflicting processes. Although, once acquired, knowledge about problems is helpful in order to achieve goal states (e.g., Funke, 1992, on complex problem solving), at earlier stages of solving an unfamiliar problem, knowledge acquisition through (free) exploration can be considered a costly investment. Likely, the novices suffered from this dilemma to allocate their cognitive resources appropriately.

Since the verbalizations were not recorded, it remains a little speculative whether the experts' self-explanations were indeed different in quality from the presumably descriptive and less efficient self-explanations of the novices. This methodological limitation could be addressed in a replication, which could also aim for a more balanced design with equally large groups of experts and novices, randomly assigned to each verbalization condition. It would be interesting to investigate linguistic and syntactic features of the verbalizations, possibly detecting more

elaborate, well-structured phrases under self-explanation conditions, indicating effortful cognitive processing. Another future perspective would be an in-depth analysis of the solutions paths (see, e.g., Hinz, 2012), which was beyond the scope of this work.

One fundamental criticism of the ToH concerns the paradigm's artificial nature and its low complexity (cf. Dörner, 1981). Betsch et al. (2011, p. 1993) compare the ToH in cognitive psychology to the fruit fly (*Drosophila melanogaster*) in biological laboratory research. The question is: What can we learn about humans from flies? What can we learn from simple, controlled laboratory experiments with the ToH about complex real-life behavior? Against the background of the present results, one basic lesson would be that expertise matters. Gaining experience takes time and effort, but once acquired, it allows experts to have a relatively clear view of the aspects relevant to the solution of a problem. Of course, like novices, experts should be given time for reflection instead of being rushed into a critical decision. There is no doubt that problem prevention and forward planning are more advisable than frantic reactions in a crisis which has gotten out of control. The warning of Dörner et al. (1983) regarding "intellectual emergency reactions" seems still very much up to date, forty years later. Self-explanations or more creative, elaborate variations of this technique can guide experts' problem solving processes, helping to avoid purely intuitive or trial-and-error reactions and fostering mental simulations of several different—both obvious and counterintuitive—solution approaches.

Acknowledgements. I would like to thank the students from summer term 2014 for recruiting and testing participants as well as for their constructive comments on planning and pretesting the experiment.

References

- Anderson, J. R., Albert, M. V., & Fincham, J. M. (2005). Tracing problem solving in real time: fMRI analysis of the subject-paced Tower of Hanoi. *Journal of Cognitive Neuroscience*, 17(8), 1261–1274. <https://doi.org/10.1162/0898929055002427>

- Betsch, T., Funke, J., & Plessner, H. (2011). *Allgemeine Psychologie für Bachelor: Denken – Urteilen, Entscheiden, Problemlösen* [General psychology for bachelor: Thinking judgment, decision making, problem solving]. Springer. <https://doi.org/10.1007/978-3-642-12474-7>
- Blech, C. (2022). *Thinking aloud, self-explanation, and the Tower of Hanoi* [OSF Project]. <https://osf.io/kz35c/>
- Blech, C., Gaschler, R., & Bilalić, M. (2020). Why do people fail to see simple solutions? Using think-aloud protocols to uncover the mechanism behind the Einstellung (mental set) effect. *Thinking & Reasoning*, 26(4), 552–580. <https://doi.org/10.1080/013546783.2019.1685001>
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, 4(1), 55–81. [https://doi.org/10.1016/0010-0285\(73\)90004-2](https://doi.org/10.1016/0010-0285(73)90004-2)
- Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145–182. https://doi.org/10.1207/s15516709cog1302_1
- Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–79). Erlbaum.
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung* [Problem solving as information processing]. Kohlhammer.
- Dörner, D. (1981). Über die Schwierigkeiten menschlichen Umgangs mit Komplexität [The difficulties of human encounters with complexity]. *Psychologische Rundschau*, 32(3), 163–179.
- Dörner, D., Kreuzig, H. W., Reither, F., & Stäudel, T. (1983). *Lohhausen: Vom Umgang mit Unbestimmtheit und Komplexität* [Lohhausen: Problem solving in uncertain and complex problems]. Hans Huber.
- Fansher, M., Shah, P., & Hélié, S. (2022). The effect of mode of presentation on Tower of Hanoi problem solving. *Cognition*, 224, 105041. <https://doi.org/10.1016/j.cognition.2022.105041>
- Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, 137(2), 316–344. <https://doi.org/10.1037/a0021663>

- Frensch, P. A., & Funke, J. (Eds.). (1995). *Complex problem solving: The European perspective*. Lawrence Erlbaum Associates, Inc.
- Funke, J. (1983). *Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor* [Some comments on the problems in problem-solving research, or: Is test intelligence still a predictor]? *Diagnostica*, 29, 283–302.
- Funke, J. (1992). *Wissen über dynamische Systeme: Erwerb, Repräsentation und Anwendung* [Knowledge about systems: Acquisition, representation, and application]. Springer. <https://doi.org/10.1007/978-3-642-77346-4>
- Hertel, J. (2014). *FlexibleDesign* [Unpublished computer software].
- Hinz, A. M. (2012). Graph theory of tower tasks. *Behavioural Neurology*, 25(1), 13–22. <https://doi.org/10.3233/BEN-2012-0345>
- Hinz, A. M., Klavžar, S., Milutinović, U., & Petr, C. (2018). *The Tower of Hanoi. Myth and math* (2nd ed.). Birkhäuser.
- Jostmann, N. B., & Gieselmann, A. (2014). When you have to climb downhill to reach the top: The effect of action versus state orientation on solving a goal-subgoal conflict in the Tower of Hanoi task. *Experimental Psychology*, 61(5), 394–401. <https://doi.org/10.1027/1618-3169/a000260>
- Knapp, F., Viechtbauer, W., Leonhart, R., Nitschke, K., & Kaller, C. P. (2017). Planning performance in schizophrenia patients: A meta-analysis of the influence of task difficulty and clinical and sociodemographic variables. *Psychological Medicine*, 47(11), 2002–2016. <https://doi.org/10.1017/S0033291717000459>
- Luchins, A. S. (1942). Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, 54(6), i–95. <https://doi.org/10.1037/h0093502>
- Mataix-Cols, D., & Bartrés-Faz, D. (2002). Is the use of the wooden and computerized versions of the Tower of Hanoi puzzle equivalent? *Applied Neuropsychology*, 9(2), 117–120. https://doi.org/10.1207/S15324826AN0902_8
- Neuman, Y., & Schwarz, B. (1998). Is self-explanation while solving problems helpful? The case of analogical problem-solving. *British Journal of Educational Psychology*, 68(1), 15–24. <https://doi.org/10.1111/j.2044-8279.1998.tb01271.x>
- Resing, W. C. M., Vogelaar, B., & Elliott, J. G. (2020). Children’s solving of ‘Tower of Hanoi’ tasks: Dynamic testing with the help of a robot. *Educational Psychology*, 40(9), 1136–1163. <https://doi.org/10.1080/01443410.2019.1684450>

- Schooler, J. W. (2011). Introspecting in the spirit of William James: Comment on Fox, Ericsson, and Best (2011). *Psychological Bulletin*, 137(2), 345–350. <https://doi.org/10.1037/a0022390>
- Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cognitive Psychology*, 7(2), 268–288. [https://doi.org/10.1016/0010-0285\(75\)90012-2](https://doi.org/10.1016/0010-0285(75)90012-2)
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. Academic Press.
- Welsh, M. C., & Huizinga, M. (2005). Tower of Hanoi disk-transfer task: Influences of strategy knowledge and learning on performance. *Learning and Individual Differences*, 15, 283–298. <https://dx.doi.org/10.1016/j.lindif.2005.05.002>
- Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., & Kliegel, M. (2014). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology*, 50(1), 304–315. <https://doi.org/10.1037/a0032982>

Complex Problem Solving and Dynamic Decision Making: What is the Difference?

CLEOTILDE GONZALEZ

Carnegie Mellon University, USA

Abstract. Complex Problem Solving (CPS) and Dynamic Decision Making (DDM) are often used as synonymous concepts. In this paper, I will document the similarities between CPS and DDM that emerge from the environments that researchers investigate and from the methodologies they use. However, I will argue that some of the demands and required cognitive processes for general complex problem solving are different from those needed for making decisions in dynamic environments. The distinction between CPS and DDM is of practical and theoretical relevance; it matters for how researchers address the question of people's complex problem solving in dynamic environments, and it also matters to represent cognitive processes precisely in algorithms that can be translated into a computational form. Generating predictive theories of DDM may be very distinct from computational representations of CPS processes.

Herbert Simon used the metaphor of an ant making his way home in a rugged beach landscape to introduce the initial ideas of problem solving (Simon, 1969). The ant makes her way from a starting point to a goal (home or a food source) around a path that is hard to describe: it appears irregular, complex, and not a straight line. Generally, the wavy path emerges because the ant cannot foresee all the obstacles in between the starting point and the goal. Instead, the ant must

adapt its course based on the difficulties that she might encounter on the way. This metaphor was used to exemplify how humans may address complex problems, and to postulate the study of human reasoning in the information processing theory entitled *Human Problem Solving* (Newell & Simon, 1972). In this theory, a thinking human being is an *adaptive* system, and the behavior is a result of the properties of the human mind and the properties of the environment.

However, it became clear that the kind of problems that a human could address under the general problem solving idea would need to be relatively simple and well defined, in contrast to real-world problems that are ill-defined and cannot easily be translated to mathematical solutions (see Dörner & Funke, 2017, for a historical discussion). The concept of Complex Problem Solving (CPS) advanced the general notion of problem solving with the need to consider “complex” environments. In a historical review of the origins of CPS, Dörner and Funke (2017) document the introduction of the CPS concept to the mid-1970s (Dörner, 1975), and the increased popularity of CPS research after the edited volumes by Sternberg and Frensch (1991) and Frensch and Funke (1995).

The key point of CPS is the *problem complexity*, the type of ill-defined environments that have no clear goal state or well-defined means of moving towards the goal state. In a related and contemporary research tradition, Dynamic Decision Making (DDM) was defined by the characteristics of a complex and dynamic environment in which decisions are made (Brehmer, 1992; Brehmer & Dörner, 1993). DDM involves “conditions which require a series of decisions, where the decisions are not independent, where the state of the world changes, both autonomously and as a consequence of the decision maker’s actions, and where the decisions have to be made in real time” (Brehmer, 1992, p. 211). The core aspect of any decision theory is the act of *choice*, the agent’s selection among options that might be available concurrently or in a sequence (Payne et al., 1993). An essential element of DDM is the idea of making a *sequence* of decisions (in contrast to making a single choice among simultaneously presented alternatives), based on what was learned from the consequences of decisions made in the past (Edwards, 1962).

This background explains why CPS and DDM are often used as synonymous concepts (e.g., Dörner & Funke, 2017; Dörner & Güss, 2022). They both depend on ill-defined environments of similar characteristics, and they both appear to require a process by which an agent moves towards a goal by making a sequence of decisions, often based on what the agent has learned from past choices. But, are CPS and DDM synonymous? I will argue that while there are similarities between CPS and DDM, they are not synonymous. Differences between CPS and DDM take place in the theoretical developments of algorithmic representations that require significantly more precision and systematic processes. A distinction between these research areas matters for how researchers address the question of people's complex problem solving in dynamic environments and for how research on algorithmic representations will advance.

Two Determinants of Human Behavior

Perhaps the distinction between CPS and DDM can be best illustrated by highlighting the two different but interrelated determinants of human behavior: human mind and the environment. These two elements of human behavior are explained by another famous Simon's metaphor: A pair of scissors in which one blade is the human mind and the other blade is the structure of the environment (Simon, 1991). The idea is that human behavior must be understood by the *joint* effect of the two blades in the scissors. One cannot understand behavior by separately studying the human mind from the environmental structure. In fact, the idea of the scissors metaphor dates back to Simon's earlier work on rational choice and the concept of bounded rationality (Simon, 1956). In the work that led to his Nobel Prize, he introduced the "satisficing" human, who cannot optimize her decisions because the computational demands of the decision environment might be beyond her mental capabilities. Thus, a satisficing human decision maker makes choices that are "good enough" for the demands of the particular environment.

To clarify the similarities and highlight the differences between CPS and DDM, it is relevant to go back to the two blades of Simon's scissors. I will argue that the

similarities between CPS and DDM emerge from the environments that researchers of CPS and DDM investigate and from the methodologies they use. But I will also argue that their distinction relies on the demands and required mental processes which need to be made precise and systematically described to generate predictive theories of CPS and DDM.

Complex Cognition and Complex Environments

In the study of CPS, there has been confusion about where the complexity emerges from, whether it emerges from the cognition or from the environment. Complex cognition has been understood as involving high-level cognitive processes such as problem solving and decision making, as well as their interaction with basic cognitive processes including perception, learning, motivation and emotion (Schmid et al., 2011). But in many problem solving situations, as in the ant metaphor, the agent (i.e., the ant) is quite simple. Simon suggested that the apparent complexity of the agent's behavior over time is largely a reflection of the complexity of the environment, not of the ant's cognition (Simon, 1969). In fact, an organism might be simple if it has a single goal, it has a fixed aspiration level that does not change, and a definite path to achieve the goal. In contrast, complex organisms might have the capability of searching and responding to multiple goals, and satisfying the needs that reach a threshold that may change dynamically (Simon, 1957). In other words, the complexity of the cognition depends on the complexity of the environment demands.

While much effort in psychological research has been dedicated to generate integrated theories of cognition (Newell, 1994), less effort in psychology has been dedicated to create an integrated theory or a taxonomy of environments. Newell and many followers argued for the need to create a unified theory of cognition. For example, the Adaptive Control of Thought-Rational (ACT-R, Anderson & Lebiere, 1998) evolved into an integrated set of mental modules that produce a wide range of behavior including perceptual-motor modules, declarative and procedural learning. These efforts have been essential for the development of common understanding

of the mental processes involved in CPS and DDM. However, similar efforts are required for developing integrated theories of environments.

Use of Microworlds to Represent Environmental Complexity

CPS and DDM research share similarities in the investigation of complex environments and their effects on human behavior as they are studied using graphical interactive simulations, called *microworlds*. In psychological research, CPS and DDM are perhaps the areas that have contributed most insights regarding an integrated view of complex environments. Complex environments, particularly in CPS research, have often been characterized by the number of elements represented: the number of variables, the number of interconnectivities among those variables, and the way these variables and connections change over time (Schmid et al., 2011). Complex environments would correspond to a large number of variables, many interconnectivities, and complex dynamics (Dörner & Funke, 2017). For example, early CPS work involved scenarios with a large number of variables (e.g., more than 2000 variables in “Lohhausen”) in simulated microworlds. Lohhausen involved a group of activities in a small town where participants played the role of a mayor that controlled the well-being of the population (Dörner, 1980; Funke, 1988).

These highly complex scenarios are represented computationally in microworlds (Brehmer & Dörner, 1993; Gonzalez et al., 2005; Omodei & Wearing, 1995; Turkle, 1984). Microworlds represent the essential characteristics of real-world decision making and problem solving environments. However, microworlds range in the characteristics of complexity that are included. For example, *structural complexity* refers to the number of elements in the environments; number of variables, exogenous changes, and opacity, while *dynamic complexity* refers to the interrelationships of all these elements over-time (Gonzalez et al., 2005).

Microworlds used to study DDM have been characterized in a continuum of structural and dynamic complexity (Gonzalez et al., 2017). Dynamic environments may involve various degrees of change, as alternatives may vary independently

from external events or endogenously because of the decisions made previously (Edwards, 1962). Thus, DDM environments vary from structurally complex (i.e., they consist of a large number of alternatives, high time constraints, and high uncertainty) to structurally simple (i.e., they have few alternatives, no time constraints, and little uncertainty), and structurally simple tasks can have high or low dynamic complexity. Dynamic complexity emerges from the relationship between choices and their effects over time, from the sequential nature of these interdependencies, and from the various lags between actions and their effects on the environment (Gonzalez, 2017; Sterman, 1989).

Both, CPS and DDM researchers have investigated complex cognition using complex environments represented in microworlds. From this research, we have learned about extreme demands that complex environments place on human cognition and the role of experience to improve performance in such tasks (Brehmer, 1980; Diehl & Sterman, 1995; Sterman, 1994). We know that humans are generally poor at handling systems with long feedback delays (Brehmer, 1992; Sterman, 1989), and they have difficulty learning in situations involving environmental constraints, such as workload and time pressure (Gonzalez, 2004, 2005; Kerstholt & Raaijmakers, 1997).

However, recently a concrete taxonomy of human errors in CPS has revealed some of the distinctions between CPS and DDM. The errors reported along the CPS steps reflect the influence of factors beyond cognitive inability to handle complex tasks: motivation and emotions (Dörner & Güss, 2022). A taxonomy of human errors in CPS was developed using classic microworlds including: CHOCO FINE, a simulation of a chocolate production company; MORO, a simulation of a tribe with semi-nomads; and WINFIRE, a simulation of firefighting units engaging in protecting cities from wildfires (Dörner & Güss, 2022). The three microworlds differ in their characteristics of exogenous changes, number of variables, opacity, and over-time interrelationships. The observations of individuals working on these three microworlds led to the characterization of 24 errors generally grouped in categories defined by the steps required for CPS (Dörner & Güss, 2022): (1) problem identification, (2) goal definition, (3) information gathering, (4) elaboration,

tion and prediction, (5) planning, decision making and action, and (6) evaluation of outcome and self-reflection. The influence of motivational, emotional, and cognitive processes, and how they interact together has been formalized in a theory of CPS, called PSI, named as an abbreviation for the word psychology (Dörner & Güss, 2022; 2013). It is precisely in the details of this PSI theory defined for CPS, that the differences emerge from the theories of DDM.

Theories and Predictive Models

The most well-known theory of DDM is Instance-Based Learning Theory (IBLT) (Gonzalez et al., 2003). IBLT emerged from a set of behavioral phenomena discovered in experiments using microworlds, and from the efforts to implement computational algorithms that would replicate the decision process involved in the microworlds, using a cognitive architecture (Gonzalez et al., 2003; Gonzalez, 2013, 2022). ACT-R (Anderson & Lebiere, 1998) is a cognitive architecture that intends to represent mathematically and computationally, the various aspects of human mind, including perception and action, learning, memory, and other processes. To demonstrate IBLT, ACT-R was used initially as the implementation platform of a concrete predictive model of the decisions that humans took in a complex microworld, called the Water Purification Plant (WPP) (Gonzalez et al., 2003). WPP is an example of a microworld that encompasses all the characteristics of DDM: multiple sequential decisions need to be made under uncertainty, in an environment that changes exogenously, and as a result of the decisions made, and in which there is high workload and time constraints (Brehmer, 1992). As such, WPP is a task that is structurally and dynamically complex and needs to be resolved from experience and extended practice by noticing patterns of exogenous events (Gonzalez, 2004, 2005; Gonzalez et al., 2003).

The theoretical process and mechanisms proposed in IBLT have been used in the computational implementation of a large number of models. It is important to highlight that IBLT is not only a picture with a description of the processes involved in DDM, but IBLT presents a concrete algorithm and the associated mathematical

mechanisms that are used to implement the algorithm in computational form. IBLT's algorithm has been published in multiple places, but I recommend Gonzalez et al. (2003), Gonzalez and Dutt (2011), Gonzalez (2013), and to more recent presentations of the theory, Nguyen, Phan, and Gonzalez (2023) for the details.

The IBL decision process involves the following steps: (1) recognition and retrieval of past experiences (i.e., instances) according to their similarity to an ongoing decision situation, (2) generation of the expected utility of various decision alternatives by using past experiences, (3) choice of the option that best generalizes past experiences to new decisions, and (4) feedback processes that update past experiences based on the observation of decision outcomes.

In IBLT, an “instance” is a memory unit that results from the potential alternatives evaluated. These memory representations consist of three elements, which are constructed over time: a situation (a set of attributes that give a context to the decision), a decision (the action taken corresponding to an alternative in a state), and a utility (expected utility or experienced outcome of the action taken in a state). Each instance in memory has an *activation* value, which represents how readily available that information is in memory, and it is determined by similarity to past situations, recency, frequency, and noise (Anderson & Lebiere, 2008). Activation of an instance is used to determine the probability of retrieval of an instance from memory (i.e., *cognitive probability*) which is a function of its activation relative to the activation of all instances in memory. The expected utility of a choice option is calculated based on *blending* past outcomes. This blending mechanism for choice has its origins in a more general blending formulation (Lebiere, 1999), but in discrete choice models, blending is defined as the sum of all past experienced outcomes weighted by their probability of retrieval (e.g., Gonzalez & Dutt, 2011; Lejarraga et al., 2012). This formulation of blending represents the general idea of an *expected value* in decision making, where instead of using the factual probability, blending uses a cognitive probability, a function of the activation equation in ACT-R. At each time step, the IBL algorithm recognizes a situation in the environment (based on similarity), calculates the expected utility of the option being evaluated, determines when to stop evaluating additional alternatives, and at that

point decides to make a choice by selecting the option that has the highest blended value. Feedback, which might be immediate or delayed, updates all instances in memory that lead to a particular outcome in the task. This process goes on over time.

IBLT and the power of the predictions of IBL models compared to human decisions have been demonstrated in a large diversity of tasks that vary in structural and dynamic complexity. The demonstrations include structurally and dynamically simple tasks, such as binary choice (Gonzalez & Dutt, 2011; Lejarraga et al., 2012); structurally simple but dynamically complex tasks such as control of carbon dioxide in the atmosphere (Gonzalez & Dutt, 2011) and supply chain inventory management; structurally complex and dynamically simple tasks such as search and rescue and navigation in non-stochastic situations (Nguyen et al., 2023); and structurally and dynamically complex tasks, such as dynamic resource allocation of resources (Gonzalez et al., 2003) and cyber defense (Cranford et al., 2020; Aggarwal et al., 2022).

Theories of Complex Problem Solving

For CPS, a main theory is PSI (Dörner & Güss, 2013, 2022). In contrast to IBLT, PSI aims at integrating not only the cognitive processes but motivational and emotional processes as well. Because of addressing the complex challenge of integrating motivation, cognition, and emotion into a single theory, PSI is extremely complex, and it can be difficult to validate (Dörner & Güss, 2013). The validation approach of PSI has been to determine whether the theory predicts potential reactions of humans to specific conditions. Researchers have attempted to conduct experimental research to compare the behavior predicted by PSI with the actual behavior of humans. For example, researchers analyzed the strategies of 30 participants and compared their behavior with 30 different PSI agents to explore a simulated island and how they solved the problem of survival on the island. The authors compared information on “when planning occurred and to what extent, the number of locations on the island visited by participants at the beginning and at the end

of the simulation, what emotions were shown, when those emotions changed and how often” (according to Dörner & Güss, 2013, p. 312, these results are reported in Dörner et al., 2002, and in Detje, 1999, but both of these references are written in German, not translated to English). Other efforts to validate PSI involve using the steps in the theory to explain cases of complex historical events (e.g., Dörner & Güss, 2011). Using historical documents, the authors analyzed decision making errors and potential causes for those errors using PSI.

Dörner and Güss (2013) suggest that PSI theory may be compared to other theories, for example, to ACT-R. However, for such a comparison to take place, the process and mechanisms proposed in PSI will need to be converted into algorithmic and mathematical form. To make an algorithm computational, the set of steps proposed in a theory should have a level of precision and systematic calculation so that they could be carried out automatically (Chabert, 1999). Unfortunately, PSI appears to be imprecise, thus making the generation of computational algorithms impossible. Furthermore, to achieve the level of computational algorithms, the mathematical formulations of the PSI steps would need to be developed and validated against human behavior.

Conclusions

This manuscript aims at clarifying the similarities and differences between CPS and DDM. In many publications CPS and DDM are used as synonymous concepts. However, I argue that they are not synonymous and should not be used interchangeably.

In this paper, I clarify that there are important similarities between CPS and DDM. The similarities emerge from the common interest and definitions of the study of complex environments. Both CPS and DDM engage in the investigation of ill-defined environments that demand a series of interdependent decisions, made in real-time, while the state of the world changes, both autonomously and because of the decision maker’s actions (Dörner & Funke, 2017; Gonzalez et al., 2017). Furthermore, both CPS and DDM use microworlds, which are computational

interactive representations of complex tasks, to investigate human behavior in experimental studies (Brehmer, 1992; Gonzalez et al., 2005).

However, CPS and DDM differ in the demands and required cognitive processes, and in the theoretical representations of these processes. When the goal is to advance theory and to develop precise predictive computational models of behavior, the specific processes and mechanisms used in these advancements are very distinct for CPS and DDM. PSI, the relevant theory for CPS is a large-encompassing endeavor, aiming at including cognitive, motivational, and emotional processes and their interactions (Dörner & Güss, 2013, 2022). In contrast, IBLT, the relevant theory for DDM, is a cognitive theory that has little or no relevance for motivational and emotional processes (Gonzalez et al., 2003; Gonzalez, 2013). Furthermore, regarding the development of predictive behavioral models, IBLT provides clear and precise algorithmic and mathematical formulations that are needed to implement computational representations and make predictions in a large variety of tasks (e.g., Gonzalez et al., 2003; Nguyen, et al., 2023). Significant work will be required to turn PSI into an algorithmic and mathematical theory.

Author Note. Joachim Funke's work represents key intersections between Complex Problem Solving (CPS) and Dynamic Decision Making (DDM) research. I am honored to be part of this commemorative publication to celebrate Jofu's contributions, and to enhance our understanding of the similarities and distinctions between CPS and DDM. I thank all the organizations that sponsor my research and all the current and past members of the Dynamic Decision Making Laboratory (DDMLab) at Carnegie Mellon University.

References

- Aggarwal, P., Thakoor, O., Jabbari, S., Cranford, E. A., Lebiere, C., Tambe, M., & Gonzalez, C. (2022). Designing effective masking strategies for cyberdefense through human experimentation and cognitive models. *Computers & Security, 117*, 102671. <https://doi.org/10.1016/j.cose.2022.102671>

- Anderson, J. R., & Lebiere, C. J. (1998). *The atomic components of thought* (1st ed.). Psychology Press. <https://doi.org/10.4324/9781315805696>
- Brehmer, B. (1980). In one word: Not from experience. *Acta Psychologica*, 45(1–3), 223–241. [https://doi.org/10.1016/0001-6918\(80\)90034-7](https://doi.org/10.1016/0001-6918(80)90034-7)
- Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, 81(3), 211–241. [https://doi.org/10.1016/0001-6918\(92\)90019-A](https://doi.org/10.1016/0001-6918(92)90019-A)
- Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, 9(2–3), 171–184. [https://doi.org/10.1016/0747-5632\(93\)90005-D](https://doi.org/10.1016/0747-5632(93)90005-D)
- Chabert, J. L. (1999). Euclid's Algorithm. In J. L. Chabert, *A history of algorithms: From the pebble to the microchip* (pp. 113–138). Springer. <https://doi.org/10.1007/978-3-642-18192-4>
- Cranford, E. A., Gonzalez, C., Aggarwal, P., Cooney, S., Tambe, M., & Lebiere, C. (2020). Toward personalized deceptive signaling for cyber defense using cognitive models. *Topics in Cognitive Science*, 12(3), 992–1011. <https://doi.org/10.1111/tops.12513>
- Detje, F. (1999). *Handeln erklären*. Deutscher Universitätsverlag.
- Diehl, E., & Sterman, J. D. (1995). Effects of feedback complexity on dynamic decision making. *Organizational Behavior and Human Decision Processes*, 62(2), 198–215. <https://doi.org/10.1006/obhd.1995.1043>
- Dörner, D. (1975). Wie Menschen eine Welt verbessern wollen und sie dabei zerstören. *Bild der Wissenschaft*, 12(2).
- Dörner, D. (1980). On the difficulties people have in dealing with complexity. *Simulation & Games*, 11(1), 87–106. <https://doi.org/10.1177/104687818001100108>
- Dörner, D. (2002). *Die Mechanik des Seelenwagens: Eine neuronale Theorie der Handlungsregulation*. Huber.
- Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8, 1153. <https://doi.org/10.3389/fpsyg.2017.01153>
- Dörner, D., & Güss, C. D. (2011). A psychological analysis of Adolf Hitler's decision making as commander in chief: Summa confidentia et nimius metus. *Review of General Psychology*, 15(1), 37–49. <https://doi.org/10.1037/a0022375>

- Dörner, D., & Güss, C. D. (2013). PSI: A computational architecture of cognition, motivation, and emotion. *Review of General Psychology*, 17(3), 297–317. <https://doi.org/10.1037/a0032947>
- Dörner, D., & Güss, C. D. (2022). Human error in complex problem solving and dynamic decision making: A taxonomy of 24 errors and a theory. *Computers in Human Behavior Reports*, 7, 100222. <https://doi.org/10.1016/j.chbr.2022.100222>
- Dutt, V., & Gonzalez, C. (2012). Human control of climate change. *Climatic Change*, 111(3), 497–518. <https://doi.org/10.1007/s10584-011-0202-x>
- Edwards, W. (1962). Dynamic decision theory and probabilistic information processings. *Human Factors*, 4(2), 59–74. <https://doi.org/10.1177/001872086200400201>
- Frensch, P. A., & Funke, J. (Eds.) (1995). *Complex problem solving: The European perspective*. Lawrence Erlbaum Associates.
- Funke, J. (1988). Using simulation to study complex problem solving: A review of studies in the FRG. *Simulation & Games*, 19(3), 277–303. <https://doi.org/10.1177/0037550088193003>
- Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors*, 46(3), 449–460. <https://doi.org/10.1518/hfes.46.3.449.50395>
- Gonzalez, C. (2005). Task workload and cognitive abilities in dynamic decision making. *Human Factors*, 47(1), 92–101. <https://doi.org/10.1518/0018720053653767>
- Gonzalez, C. (2013). The boundaries of Instance-Based Learning Theory for explaining decisions from experience. *Progress in Brain Research*, 202, 73–98. <https://doi.org/10.1016/B978-0-444-62604-2.00005-8>
- Gonzalez, C. (2017). Decision making: A cognitive science perspective. In S. E. F. Chipman (Ed.), *The Oxford Handbook of Cognitive Science* (pp. 249–263). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199842193.001.0001>
- Gonzalez, C. (2022). Learning and dynamic decision making. *Topics in Cognitive Science*, 14(1), 14–30. <https://doi.org/10.1111/tops.12581>
- Gonzalez, C., & Dutt, V. (2011). Instance-based learning: Integrating decisions from experience in sampling and repeated choice paradigms. *Psychological Review*, 118(4), 523–551. <https://doi.org/10.1037/a0024558>

- Gonzalez, C., Fakhari, P., & Busemeyer, J. (2017). Dynamic decision making: Learning processes and new research directions. *Human Factors*, 59(5), 713–721. <https://doi.org/10.1177/0018720817710347>
- Gonzalez, C., Lerch, F. J., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, 27, 591–635. https://onlinelibrary.wiley.com/doi/abs/10.1207/s15516709cog2704_2
- Kerstholt, J. H., & Raaijmakers, J. G. (1997). Decision making in dynamic task environments. In R. Crozier, J. G. Johnson, & O. Svenson, *Decision making: Cognitive models and explanations* (pp. 205–217). Routledge.
- Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft*, 8(1), 5–19. <https://doi.org/10.1007/BF03354932>
- Lejarraga, T., Dutt, V., & Gonzalez, C. (2012). Instance-based learning: A general model of repeated binary choice. *Journal of Behavioral Decision Making*, 25(2), 143–153. <https://doi.org/10.1002/bdm.722>
- Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
- Nguyen, T. N., Phan, D. N., & Gonzalez, C. (2023). SpeedyIBL: A comprehensive, precise, and fast implementation of instance-based learning theory. *Behavior Research Methods*, 55, 1734–1757. <https://doi.org/10.3758/s13428-022-01848-x>
- Omodei, M. M., & Wearing, A. J. (1995). The Fire Chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior Research Methods, Instruments, & Computers*, 27(3), 303–316. <https://doi.org/10.3758/BF03200423>
- Payne, J. W., Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.
- Schmid, U., Ragni, M., Gonzalez, C., & Funke, J. (2011). The challenge of complexity for cognitive systems. *Cognitive Systems Research*, 12(3–4), 211–218. <https://doi.org/10.1016/j.cogsys.2010.12.007>
- Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129–138. <https://psycnet.apa.org/doi/10.1037/h0042769>
- Simon, H. A. (1957). *Models of man; social and rational*. Wiley.
- Simon, H. A. (1969). *The sciences of the artificial*. MIT Press.

- Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization Science*, 2(1), 125–134. <https://doi.org/10.1287/orsc.2.1.125>
- Sternberg, R. J., & Frensch, P. A. (Eds.). (1991). *Complex problem solving: Principles and mechanisms*. Lawrence Erlbaum Associates. <https://doi.org/10.4324/9781315807546>
- Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, 35(3), 321–339. <https://doi.org/10.1287/mnsc.35.3.321>
- Sterman, J. D. (1994). Learning in and about complex systems. *System Dynamics Review*, 10(2–3), 291–330. <https://doi.org/10.1002/sdr.4260100214>
- Turkle, S. (1984). *The second self: Computers and the human spirit*. MIT Press.

Psychologische Theorien als finite Automaten

WOLFGANG SCHOPPEK

Universität Bayreuth

Abstract. Many psychological theories can be described as specifying psychic states and conditions that govern transitions among those states. These theories' structures resemble that of a finite automaton, specifically a non-deterministic one where state transitions occur with certain probabilities. The chapter explores the potential of using automata theory to reconstruct psychological theories, using a deterministic automaton based on the Moore model for simplicity. This is demonstrated with the example of the strength model of self-control and applied to a question from problem solving research, asking for the conditions that trigger effortful reasoning. Such an endeavor encourages a systematic reflection on the states in which a psychological system may exist. By considering how many dimensions these states can be described with and whether these dimensions must be independent of each other, it prompts deeper thought. Since states play a role in many theories, comparing them allows for potential unification by identifying equivalent states. Exploring the relevant dimensions for describing states reveals that psychological theories often address only very specific states. This raises questions about the significance of neglected states and the potential for new discoveries within them. Ultimately, the question arises as to whether finite automata can be an appropriate model for the human psyche. While models are simplified representations of reality, they need not equate to viewing the human psyche itself as a finite automaton. It is possible to analyze processes mechanistically without questioning the intentionality of the acting subject.

Joachim Funke und mich verbindet seit vielen Jahren die Überzeugung, dass in der Psychologie zu wenig Auseinandersetzung auf der Ebene von Theorien stattfindet. Das zeigt sich daran, dass es nur wenige Theorien gibt, die diesen Namen auch verdienen, dass es in Publikationen als unschicklich gilt, Arbeiten zu zitieren, die älter als zehn Jahre sind, oder auch daran, dass die theoretische Grundlage vieler Forschungsarbeiten als dünn bezeichnet werden muss. Jüngsten Ausdruck fand diese Haltung in dem Buch *Wohin steuert die Psychologie?*, das Alexander Wendt und Joachim Funke 2022 vorgelegt haben (Wendt & Funke, 2022). Viele weitere Autoren teilen unsere Einschätzung und belegen diese mit empirisch-bibliometrischen Studien und konzeptuellen Arbeiten. So fanden McPhetres und Kollegen (2021), dass von 2009 bis 2019 in *Psychological Science* in nur rund 15 % der Artikel explizit berichtet wurde, dass Vorhersagen einer Theorie getestet wurden. Auch Wendt und Wolfradt (2022) konstatieren in einer historischen Analyse der Zeitschrift *Psychological Research* seit 1922 einen Rückgang theoretischer Positionierungen einzelner Autoren zugunsten kürzerer empirischer Arbeiten mit mehreren Autoren. Auf konzeptueller Ebene argumentieren Borgstede und Eggert (2023), dass die verbreitete Methodik der Pfadanalyse latenter Variablen ohne zugrunde liegende Theorien kaum bedeutsame und replizierbare Ergebnisse zu erzeugen vermag (siehe auch Oberauer & Lewandowsky, 2019).

In meinem Beitrag stelle ich eine Idee zum Umgang mit Theorien vor, die eine neue Auseinandersetzung anregen soll. Sie besteht darin, die in Theorien angenommenen Zustände und deren Übergänge als finite Automaten zu rekonstruieren. Gezeigt wird das zunächst am Beispiel einer Selbstregulationstheorie, dann anhand einer Fragestellung aus der Problemlöseforschung. Abschließend wird diskutiert, welches Potenzial ein solcher Ansatz bietet, aber auch an welche Grenzen er stößt.

Viele psychologische Theorien können strukturell so beschrieben werden, dass sie unterscheidbare Zustände des Individuums annehmen und Aussagen über die Bedingungen und Wahrscheinlichkeiten des Übergangs von einem Zustand in den anderen machen. So werden in der Flow-Theorie (Csikszentmihalyi et al., 2014) vier Zustände mit charakteristischen Auswirkungen auf das Erleben definiert.

Am prominentesten ist natürlich der Flow-Zustand, der dadurch gekennzeichnet ist, dass Handelnde ganz in ihrem Handeln aufgehen (keine Selbstbeobachtung, verändertes Zeitgefühl, Gefühl der Leichtigkeit, trotz anspruchsvoller Tätigkeit). Daneben gibt es die Zustände der Apathie, der Langeweile und der Überforderung, letztere verbunden mit Angst. Der Übergang in den Flow-Zustand wird dadurch gefördert, dass die Situation hohe Anforderungen stellt, denen das Individuum mit hoch ausgeprägten Fertigkeiten begegnet, sodass ein Gleichgewicht von Anforderungen und Fertigkeiten auf hohem Niveau entsteht. Weitere Beispiele sind die Rubikon-Theorie mit ihren Übergängen zwischen Zuständen der Realitätsorientierung und der Realisierungsorientierung (Gollwitzer, 1990) oder das *Strength Model of Self-Regulation* (Baumeister et al., 2007), was hier als Willenskraft-Theorie übersetzt wird, mit Zuständen, die durch unterschiedliche Anwendung und Verfügbarkeit von Willenskraft gekennzeichnet sind.¹ Die Annahme von diskreten Zuständen ist in diesen Fällen aus meiner Sicht genau so gemeint und nicht als eine künstliche, vereinfachende Diskretisierung zu verstehen.

Die Struktur dieser Theorien erinnert an einen finiten Automaten – in der Spielart eines nicht-deterministischen Automaten, bei dem die Zustandsübergänge mit bestimmten Wahrscheinlichkeiten auftreten. Im vorliegenden Beitrag unternehme ich den Versuch, diese Ähnlichkeit wörtlich zu nehmen und das Potenzial einer automatentheoretischen (Re-)Konstruktion psychologischer Theorien auszuloten. Der Einfachheit halber bediene ich mich hierbei eines deterministischen Automaten nach dem Moore-Modell (Moore, 1956).

¹ Mir ist bewusst, dass die Auffassung von Willenskraft als begrenzte Ressource im Zuge der Replikationskrise angezweifelt wird (z. B. Inzlicht & Berkman, 2015). Aber erstens muss man die Frage danach laut Friese et al. (2019) als nicht geklärt betrachten; zweitens funktioniert die Willenskraft-Hypothese als Beispiel für meinen Ansatz unabhängig davon, für wie sicher ihre empirische Bestätigung angesehen wird.

Formale Überlegungen

Ein Moore-Automat kann als ein 6-Tupel $A = (Q, \Sigma, \Omega, \delta, \lambda, q_0)$ definiert werden, dessen Elemente folgende Bedeutung haben:

Q : endliche Menge von Zuständen
 Σ : Eingabealphabet
 Ω : Ausgabealphabet
 δ : Übergangsfunktion: $\delta: Q \times \Sigma \rightarrow Q$
 λ : Ausgabefunktion: $\lambda: Q \rightarrow \Omega$
 q_0 : Anfangszustand

Die Menge von Zuständen Q besteht bei der Rekonstruktion psychologischer Theorien aus unterscheidbaren psychischen Zuständen. Das Eingabealphabet Σ besteht aus externen Ereignissen, das Ausgabealphabet Ω aus Verhalten. Die Übergangsfunktion δ beschreibt die Regeln, wie sich der Zustand des Automaten in Abhängigkeit vom aktuellen Zustand und der aktuellen Eingabe verändert. Dabei gibt es für jede Kombination aus den Elementen von Q und Σ (Cartesisches Produkt) eine Regel. Die Ausgabefunktion λ gibt für jeden Zustand ein zugehöriges Verhalten an. Ein Element aus Q wird als der Anfangszustand (q_0) definiert.

In den hier beschriebenen Rekonstruktionen werden psychische Zustände mithilfe von Zustandsdimensionen auf einem groben Auflösungsniveau definiert. Feine Abstufungen innerhalb der Zustände wie unterschiedliche Inhalte von Gedanken oder Wahrnehmungen werden nicht abgebildet. Auch Ereignisse (Elemente des Eingabealphabets Σ) werden teilweise anhand mehrerer Dimensionen klassifiziert. Die Dimensionen werden zunächst als voneinander unabhängig angesehen – eine Annahme, die in konkreten Fällen geprüft werden muss.

Aus diesen – hier noch sehr abstrakten – Beschreibungen ergibt sich zum Beispiel, dass es bei drei Zustandsdimensionen zu je zwei Abstufungen $2^3 = 8$ unterscheidbare Zustände, bei zwei Ereignisdimensionen zu je drei Abstufungen $3^2 = 9$ mögliche Ereignisse gibt, was bedeutet, dass δ aus $8 \times 9 = 72$ Übergangsregeln besteht.

Allgemein gilt für die Anzahl von Übergangsregeln R :

$$R = f^E t^S \quad (1)$$

R : Anzahl von Regeln

E : Anzahl von Ereignisdimensionen

f : Anzahl der Stufen von Ereignisdimensionen

S : Anzahl von Zustandsdimensionen

t : Anzahl der Stufen von Zustandsdimensionen

Man sieht hier, dass selbst bei einer sehr holzschnittartigen Modellierung psychischer Zustände und Ereignisse eine große Anzahl von unterschiedlichen Übergangsregeln denkbar ist – was die Realität der meisten gängigen psychologischen Theorien offensichtlich nicht widerspiegelt. Das folgende ausführlichere Beispiel wird darüber Aufschluss geben, woran das liegen könnte.

Beispiel Selbstregulation

Das folgende Beispiel analysiert die Willenskraft-Theorie von Baumeister et al. (2007). Wie bereits erwähnt, bezeichnen die Autoren selbst ihre Idee als Modell, was ich zum Anlass nehme, kurz über die Verwendung der Begriffe *Theorie* und *Modell* zu reflektieren. In der psychologischen Fachliteratur wird oft nicht scharf zwischen den Begriffen unterschieden beziehungsweise werden sie uneinheitlich verwendet. In vielen Fällen verwenden Autoren *Modell* wohl, um die anspruchsvolle, hohe Erwartungen weckende Bezeichnung *Theorie* zu vermeiden. Ein Modell wäre in diesem Sinne dann so etwas wie eine „kleine Theorie“ oder eine Theorie mit einem eher kleinen Anwendungsbereich. Anstelle dieser wenig trennscharfen Unterscheidung verwende ich den Begriff *Modell* für eine konkrete Darstellung, Umsetzung oder Rekonstruktion einer Theorie. So basieren kognitive Modelle zwar normalerweise auf Theorien wie beispielsweise ACT-R (Anderson, 2007), enthalten aber zusätzliche Annahmen, die nötig sind, um das Modell lauffähig zu machen, und nicht von der Theorie gedeckt werden. Das zeigt sich auch im Folgenden, weshalb ich die zu beschreibenden automatentheoretischen Rekonstruktionen auch als Modelle bezeichne.

Das Willenskraft-Modell (Baumeister et al., 2007) geht im Kern davon aus, dass die Aufwendung von Selbstregulation, die als Fähigkeit von Organismen definiert ist, ihre Reaktionen zu modifizieren (Baumeister & Vohs, 2007), auf eine begrenzte Ressource zurückgreift. Ist diese Willenskraft erschöpft, gelangt die Person in den Zustand der *Ego-Depletion* (ED), der dadurch gekennzeichnet ist, dass kaum mehr Selbstregulation oder Selbstkontrolle (SK) stattfindet. Diese Annahmen können mithilfe der folgenden Zustands- und Ereignis-Dimensionen dargestellt werden.

S_1 {Ego Depletion, keine ED}	E vs. e
S_2 {aktuelle Aufwendung von Selbstkontrolle, keine SK}	S vs. s
S_3 {vorherige Aufwendung von Selbstkontrolle, keine SK}	P vs. p
E_1 {SK gefordert, SK nicht gefordert}	K vs. k

$$R_{max} = 16$$

Ein Zustand ergibt sich dabei aus einer Kombination von Ausprägungen auf den drei Zustandsdimensionen. Die Ereignismenge bildet sich aus einer Dimension mit zwei Ausprägungen und besteht damit aus nur zwei Elementen.

Die Ausgabefunktion bleibt zunächst offen, da sich darüber kaum Aussagen in der Theorie finden. Folgende Regeln sind aus der Literatur zu entnehmen (dargestellt sind die Ausprägungen von S_1 bis S_3):

- a) $esp + K \rightarrow eSp$
- b) $eSp + K \rightarrow eSP$
- c) $eSP + K \rightarrow ESP$
- d) $ESP + K \rightarrow EsP$
- e) $Esp + k \rightarrow esp$

- Zu a) esp bezeichnet einen ausgeruhten Ruhezustand; wenn SK gefordert ist, wird sie aufgewendet.

- Zu b) eSp ist der Zustand, in dem erfolgreich die geforderte SK angewendet wird; das geht eine Zeit lang gut, resultiert aber im Zustand eSP , der anzeigt, dass schon länger SK angewendet wurde.
- Zu c) Bei weiter geforderter SK kommt es zur Ego Depletion; ich nehme hier an, dass mit deren Auftreten die SK nicht sofort aufhört, sondern erst im nächsten Schritt (d).
- Zu e) Der zeitliche Verlauf der Erholung von Ego Depletion ist in den Darstellungen der Theorie nicht genauer spezifiziert; gezeigt ist hier ein Zustand, bei dem schon eine Zeit lang keine SK mehr angewendet wurde.

Man sieht, dass diese Regeln nicht einmal die Hälfte der in diesem Modell nötigen Übergangsregeln abdecken. Im Automatenmodell muss nämlich für jeden möglichen Zustand definiert werden, was beim Auftreten der jeweiligen Ereignisse passiert (selbst wenn nichts passiert). Was ist also mit den restlichen Regeln? Die meisten davon ergeben sich daraus, dass implizite Annahmen der Theorie auf mehrere Zustände anwendbar sind. Konkret sind das folgende Regeln; ein Asterisk steht für eine beliebige Ausprägung der jeweiligen Zustandsdimension:

- f) Wenn keine SK gefordert wird, wird auch keine SK angewendet
($*** + k \rightarrow *_S^*$).
- g) Wenn SK gefordert wird, wird sie angewendet, sofern keine Ego Depletion vorliegt
($e^{**} + K \rightarrow eS^*$).
- h) Ego Depletion geht zurück, wenn (länger) keine SK gefordert wird
(z. B. $Esp + k \rightarrow esp$ oder $EsP + k \rightarrow Esp$).

An den Asterisken erkennt man die Anzahl der betroffenen Zustände. Für Regel f sind das acht, für Regel g sind es vier. Die Globalregeln f und g enthalten einige der Einzelregeln a bis e. Etwas komplizierter ist die Sache bei Regel h. Hier sind verschiedene Realisationen möglich. Die Abbildung 1 zeigt einen Vorschlag, wie

man den gesamten Zustands-Übergangs-Raum modellieren könnte, einschließlich des Rückgangs der Ego Depletion.

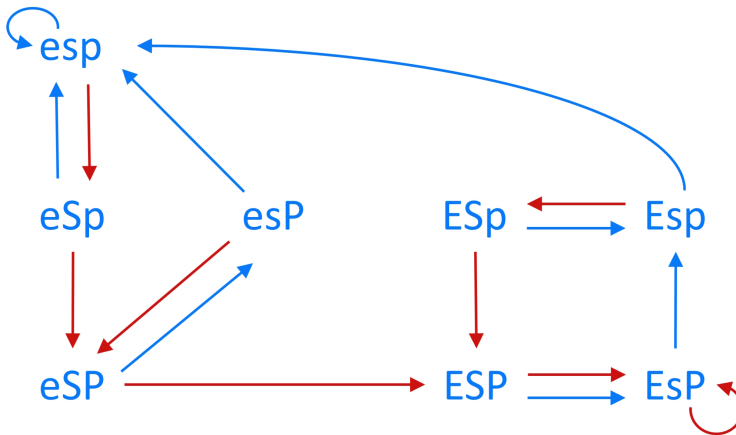


Abbildung 1. Automaten-theoretischer Rekonstruktionsversuch der Willenskraft-Theorie von Baumeister und Kollegen (2007). Rote Pfeile stehen für die Anforderung, Selbstkontrolle auszuüben, blaue Pfeile für das Fehlen solcher Anforderungen. E/e: Ego Depletion/keine Ego Depletion; S/s: Aufwendung bzw. keine Aufwendung von Selbstkontrolle; P/p: dasselbe für kürzlich erfolgte Aufwendung von Selbstkontrolle.

Was zeigt diese Analyse? Erstens, dass Theorien häufig nicht alle im Rahmen ihrer Elemente möglichen Zustände explizit ansprechen bzw. Regeln dafür formulieren. Vieles bleibt implizit und kann zwar häufig mit plausiblen Annahmen ergänzt werden, wie im Beispiel von Regel f. Das muss aber nicht immer der Fall sein, wie etwa bei der Modellierung der Erholung von ego depletion. Zweitens macht die Analyse deutlich, dass die Leistung in der Theorie weitgehend implizit bleibt. Da sie eine wichtige Rolle als Indikator für Ego Depletion spielt, kann sie aber nicht einfach ignoriert werden. Eine einfache Lösung könnte darin bestehen, dass man eine Ausgabefunktion definiert, die bei Aufwendung von SK eine höhere Leistung vorhersagt als bei Nicht-Aufwendung von SK. Das ist zwar als Verallgemeinerung zweifelhaft, würde aber für die typischen Experimente zu Ego Depletion zutreffen. Drittens ist festzustellen, dass die Angaben der Theorie zum Ablauf oder Einfluss

der Zeit ebenfalls so ungenau sind, dass zusätzliche Annahmen nötig sind. Im Beispiel ist das mit der Puffervariable S_3 sehr schlicht umgesetzt worden. Hier sind auch Lösungen mit Speicher-Variablen, die zeitlich sich verändernde Anforderungen und Aktivitäten integrieren, die bei Überschreitung von Schwellenwerten einen Zustandsübergang bewirken, denkbar. In Anbetracht der schwierigen Debatte, ob es so etwas wie ego depletion überhaupt gibt (Friese et al., 2019; Inzlicht & Berkman, 2015), ist es fraglich, ob solche detaillierten Annahmen mit Forschungen untermauert werden können.

Als Zwischenfazit ist festzuhalten, dass die Willenskraft-Theorie selbst für eine so einfache Modellierung zu unpräzise formuliert ist und dass die Modellierung wertvolle Hinweise für Präzisierungen liefert. Diese Einschätzung trifft mit ziemlicher Sicherheit auf viele Theorien zu, die rein sprachlich formuliert sind. Es ist sehr einfach, mit einem Satz eine Zusatzbedingung zu ergänzen. Im Automatenmodell wäre eine solche meist mit der Annahme einer weiteren Zustandsdimension verbunden, die die Anzahl der Zustände um den Faktor der Zahl ihrer Ausprägungen erhöht – mit den entsprechenden Konsequenzen für die Regeln der Übergangsfunktion!

Ein Beispiel zur Denkpsychologie

Wie gezeigt, kann man bestehende Theorien mit dem Instrumentarium der Automatentheorie analysieren. Man kann aber auch neue Theorien von Beginn an damit entwerfen. Wie so etwas aussehen könnte, soll in diesem Abschnitt skizziert werden. Ich wende zur Beschreibung zunächst die Theorie-Bestandteile von Funke (Wendt & Funke, 2022) an.

Die *Domäne* der Theorie-Skizze ist menschliches Denken und Handeln beim Problemlösen. Es werden Aussagen zum *Phänomen* gemacht, dass Menschen beim Problemlösen sehr unterschiedlich tief nachdenken. Bei der Beschreibung der „Tiefe des Denkens“ nehme ich Anleihen beim *Dual-Processing-Ansatz* (z. B. Evans, 2012) und definiere „tiefes Denken“ als Verarbeitung vom Typ 2. Das *Auflösungsniveau* ist dezidiert psychologisch, das heißt, der Fokus liegt auf Zu-

ständen und Vorgängen innerhalb von Individuen und die Theorieelemente sind vergleichsweise abstrakt. Es gibt keine expliziten Annahmen über die biologischen Grundlagen der unterschiedlichen Zustände (auch wenn dies eine wünschenswerte Erweiterung wäre).

Die *Theorieelemente* sind psychische Zustände, Ereignisse und Übergangsregeln. Sowohl Zustände als auch Ereignisse werden über Dimensionen definiert, die als weitgehend unabhängig angenommen werden. Wie angemessen diese Annahme ist, kann in einem ersten Schritt festgestellt werden, indem man prüft, ob alle Kombinationen von Zustandsausprägungen vorkommen können.²

Die *Struktur der Theorieelemente* entspricht in dieser Skizze einem endlichen Moore-Automaten, was eine starke Vereinfachung bedeutet, da es beim Menschen nur wenige Ereignisse gibt, die immer einen vorhersagbaren Zustandswechsel herbeiführen (wie etwa die Orientierungsreaktion). Ein nicht-deterministischer endlicher Automat dürfte zwar das angemessenere Modell sein, geht aber mit erheblichem Mehraufwand für die Ermittlung oder Schätzung der Übergangswahrscheinlichkeiten einher.

Die Zustandsdimensionen sind:

- Motivationsdifferenz Δ , definiert als Differenz zwischen den Motivationen bezüglich zweier Ziele A und B. A bezieht sich auf die Zieltätigkeit, die Nachdenken erfordert, B auf ein konkurrierendes Ziel (häufig verbunden mit Ablenkungen oder schnellen ‚Lösungen‘). Unterschieden werden die Zustände hoch ($\Delta > 0 : M$) und niedrig ($\Delta \leq 0 : m$).
- Informationsverarbeitung A , definiert nach den Konventionen des Dual-Processing-Ansatzes. Unterschieden wird zwischen Verarbeitung, die überwiegend Typ 1 zugeordnet ist, und Verarbeitung, die überwiegend Typ 2

² Man könnte die schwierige Frage nach der Abhängigkeit von Zustandsdimensionen für einen Nachteil dieser Art von Theoriekonzeption halten. Ich behaupte aber, dass diese Frage in vielen existierenden Theorien aufgrund der fehlenden Formalisierung gar nicht erst gestellt wurde. Demnach ist es ein Vorzug des hier vorgestellten Ansatzes, diese Problematik sichtbar zu machen und die Suche nach Erklärungen für Abhängigkeiten anzuregen.

zugeordnet ist (1 vs. 2). Verarbeitung vom Typ 2 ist durch eine starke Inanspruchnahme des Arbeitsgedächtnisses definiert (Evans, 2012); Verarbeitung vom Typ 1 durch ihre Automatisiertheit, das heißt, ihre Ergebnisse werden ohne erhöhte Aufmerksamkeit bereitgestellt (z. B. Erkennung eines bekannten Gesichts).

- Fertigkeiten Φ bezüglich Ziel A (niedrig: f vs. hoch: F).
- Als zusätzlicher Zustand außerhalb der Dimensionen wird das Beenden der Verfolgung von Ziel A, meist verbunden mit dem Wechsel zu Ziel B, angenommen. Diesen Zustand bezeichne ich mit E .

In einer elaborierteren Fassung der Theorie müsste als weitere Dimension noch die Motivationssumme hinzukommen, weil eine Motivationsdifferenz von 0 unterschiedliche Zustände beschreibt, je nachdem, ob die Motivationen bezüglich Ziel A und B beide niedrig oder beide hoch sind. Sinnvoll wären auch die Dimensionen Wachheit (niedrig vs. hoch), emotionale Zustände (ähnlich wie bei der Flow-Theorie) oder auch die Ego Depletion. Allerdings gewinnt bei zunehmender Anzahl von Zustandsdimensionen die schwierige Frage nach deren Unabhängigkeit an Bedeutung.³

Die Ereignisdimensionen sind:

- Anforderungen (niedrig: a vs. hoch: A) bezüglich Ziel A. Hierbei ist wichtig, dass Anforderungen möglichst unabhängig von den Fertigkeiten definiert werden. Die Anleihen bei der Flow-Theorie sind an dieser Stelle beabsichtigt.
- Selbstrelevanz (niedrig: s vs. hoch: S), ebenfalls bezüglich Ziel A. Diese wichtige Dimension ist insofern problematisch, als sie nicht unabhängig

³ Ohne hier näher darauf eingehen zu können, vermute ich eine Abhängigkeit der Dimensionen Motivationsdifferenz und Emotion. So geht Angst grundsätzlich mit erhöhter Motivation einher – mit Auswirkungen auf Δ . Motivation und Wachheit können dagegen in bestimmten Grenzen unabhängig variieren, wie Baumeister und Vohs (2007) dargelegt haben.

von der betrachteten Person bestimmt werden kann. Zum Beispiel hat die Analyse eines Harmonieverlaufs in einem Musikstück keine Selbstrelevanz für eine Person, die sich kaum für Musik interessiert. Diese Abhängigkeit führt dazu, dass einige der postulierten Übergangsregeln möglicherweise keinen empirischen Gehalt haben, sondern a priori wahr sind. Ich werde auf diesen Umstand zurückkommen.

Die wichtigsten Übergangsregeln seien zunächst sprachlich formuliert:

- a) Hohe Selbstwertrelevanz geht mit höherer Motivationsdifferenz einher.
- b) Eine hohe Motivationsdifferenz steigert die Bereitschaft zu Typ-2-Verarbeitung.
- c) Hohe Anforderungen führen bei hohen Fertigkeiten zu höherer Bereitschaft zu Typ-2-Verarbeitung.
- d) Wenn hohe Anforderungen auf niedrige Fertigkeiten stoßen, hängt der Zustandsübergang von der Selbstwertrelevanz ab. Ist diese niedrig, kommt es zu Resignation (oft mit Scheinaktivität vom Typ 1).
- e) Hohe Anforderungen, niedrige Fertigkeiten und hohe Selbstrelevanz: Üblicherweise wird für diese Kombination Angst als Konsequenz angenommen, zum Beispiel in der Flow-Theorie. Angst ist in der vorliegenden Skizze als Zustand nicht vorgesehen.
- f) Niedrige Motivationsdifferenz zusammen mit niedrigen Anforderungen und Selbstrelevanz führt zur Einstellung der Verfolgung von Ziel A.

In den formalen Darstellungen der Tabelle 1 sind die Zustände in der Reihenfolge Motivationsdifferenz Δ , Informationsverarbeitung I und Fertigkeiten Φ aufgeführt.

Aus Tabelle 1 wird ersichtlich, dass diese Übergangsregeln, die ich aus meinen Untersuchungen und Beobachtungen von Versuchspersonen beim komplexen Problemlösen gewonnen habe, für mehrere Elemente der Zustandsmenge gelten.

Tabelle 1. *Übergangsregeln der Theorieskizze zur Denkpsychologie*

Regel	Ausgangszustand	Ereignis	Folgezustand
a)	***	$*S$	$\rightarrow M^{**}$
b)	M^{**}	$**$	$\rightarrow M2^{*}$
c)	$**F$	A^{*}	$\rightarrow *2F$
d)	$**f$	As	$\rightarrow *If$
e)	$**f$	AS	$\rightarrow **f(+Angst)$
f)	m^{**}	As	$\rightarrow E$

Das hat die Theorieskizze mit der Analyse der Willenskraft-Theorie gemeinsam. Es zeigt sich also wiederholt, dass sprachlich formulierte Hypothesen oft weniger differenziert sind, als es die Annahme von Zustandsdimensionen erlauben würde. Um dieses Potenzial des Ansatzes auszuloten, sollen noch ein paar spekulative Übergangsregeln spezifischer Art betrachtet werden.

Es ist wahrscheinlich, dass eine längere tiefgehende Beschäftigung einer Person von niedrigen Fertigkeiten mit einem relevanten und schwierigen Problem zu einer Verbesserung der Fertigkeiten führt – formal ausgedrückt als $M2f + AS \rightarrow M2F$. Um diese Annahme realistisch abzubilden, müsste das Modell um eine zeitliche Komponente erweitert werden.

Eine weitere Spekulation betrifft die Frage, inwiefern die Aufwendung von Willenskraft (sensu Baumeister, 1998) bzw. die Aufwendung volitionaler Strategien mit einer vorwiegenden Typ-2-Verarbeitung gleichzusetzen ist. Die Überschneidung zwischen dem Konzept der exekutiven Funktionen (Diamond, 2013) und Typ-2-Verarbeitung (Evans, 2012) über das Arbeitsgedächtnis spricht dafür. Man könnte also annehmen, dass $mI^{*} + As \rightarrow m2^{*}$. Dieser Fall ist allerdings nur bei einer hohen Motivationssumme plausibel (die ich der Übersichtlichkeit halber im Modell weggelassen habe). Bei einer niedrigen Motivationssumme würde man eher ein Aufgeben erwarten, also $mI^{*} + As \rightarrow E$. Interessant ist in diesem Zu-

sammenhang, ob die volitionalen Strategien dann auch zu einer Erhöhung der Motivationsdifferenz führen, was dem Kern der Ach-Lewin-Kontroverse entspricht (Ach, 1910; Lewin, 1926).

Probleme der automatentheoretischen Theorie-Konzeption

Bevor ich abschließend die Vorteile des Ansatzes zusammenfasse, möchte ich systematisch auf einige Probleme eingehen, die zum Teil in den bisherigen Ausführungen bereits aufgetreten sind.

Entlehnte Konzepte: Die aus anderen Theorien entlehnten Konzepte müssten genauer spezifiziert werden. Deutlich wird das am Beispiel der Motivationsdifferenz. Für diese Zustandsdimension bin ich davon ausgegangen, dass klar ist, was mit Motivation gemeint ist. Das ist weder selbstverständlich noch trivial. Bei einer gründlichen Theoriekonzeption müssen solche Entlehnungen explizit gemacht werden. Die sich daraus ergebende Verschachtelung von Theorien kann insofern problematisch sein, als die entlehnte Theorie nicht einfach als Modul mit einer einfachen Schnittstelle eingebaut werden kann. Deutlich wird das am Beispiel der Motivationsdifferenz: Es ist offensichtlich, dass die Ereignisdimension „Anforderungen“ nicht nur im dargestellten Zustandsraum verarbeitet wird, sondern zusätzlich im Theoriemodul „Motivation“. Auch lassen sich Theorien nur dann sinnvoll integrieren, wenn sie ähnliche bzw. kompatible epistemologische Grundpositionen haben (Fahrenberg, 2015).

Die *Zeit* ist im Automaten-Modell schwierig zu berücksichtigen, da dieses eine diskrete Taktung vorsieht, deren Frequenz spezifiziert werden müsste, und zwar so, dass sie auf unterschiedliche Prozesse passt. Die im Beispiel der Selbstregulationstheorie gewählte Modellierung mit einer Puffervariablen stellt eine sehr grobe Lösung dar. Feiner könnte das in einem nicht-deterministischen Automaten über Hilfsvariablen gelöst werden, die einen zeitlichen Verlauf integrieren und ihrerseits die Übergangswahrscheinlichkeiten verändern. So würde die Übergangswahrscheinlichkeit in einem Zustand der Ego Depletion proportional zu einer Variablen steigen, die die aufgewendete Selbstregulation über die Zeit integriert.

Die Problematik der unterspezifizierten zeitlichen Dynamik psychischer Prozesse betrifft allerdings viele Theorien, nicht nur wenn sie automatentheoretisch formalisiert werden.

Auswahl und Unabhängigkeit von Zustandsdimensionen: Viele und fein abgestufte Zustandsdimensionen führen schnell zu einer unübersichtlichen Anzahl unterscheidbarer Zustände mit einer noch größeren Anzahl potenzieller Übergangsregeln. Dieses Problem kann gleichzeitig als Stärke des Ansatzes gesehen werden, weil es Autoren dazu anregt, genauer über die Zustände nachzudenken. So deuten etwa Zustände, die durch eine unmöglich erscheinende Kombination von Ausprägungen der Dimensionen gekennzeichnet sind, auf eine Abhängigkeit der Dimensionen hin (was die Zahl der möglichen Zustände wieder reduziert). Eine Lösungsmöglichkeit bei unabhängigen Dimensionen besteht darin, bestimmte Zustände explizit aus dem Anwendungsbereich der Theorie auszuschließen.

Ontologischer Status von Zuständen: Darüber hinaus stellt sich die Frage, inwieweit die Annahme diskreter Zustände eine reale Grundlage hat oder nur eine künstliche Aufteilung darstellt. Für eine reale Grundlage sprechen Erkenntnisse aus der Theorie nicht linearer dynamischer Systeme (Haken & Schiepek, 2006), die sich in unterschiedlichen, jeweils relativ stabilen Attraktoren befinden können, wobei als Ordnungsübergänge bezeichnete Wechsel von einem Attraktor zum anderen durch die Veränderung eines (oder mehrerer) Kontrollparameter ausgelöst werden. Eine solche Sichtweise findet derzeit erfolgreiche Anwendung in der klinischen Psychologie (z. B. Schiepek et al., 2014). Auch eine künstliche Diskretisierung wäre begründbar. Wichtig ist nur, die Unterscheidung explizit zu machen. In den hier skizzierten Beispielen halte ich Flow für einen Zustand, der im Sinne eines eigenen Attraktors von anderen Zuständen abgrenzbar ist. Man kann nicht „ein bisschen im Flow“ sein (wer das annimmt, hat wahrscheinlich noch keinen echten Flow-Zustand erlebt). Dahingegen sind die im Selbstregulationsmodell angenommenen Abstufungs-Zustände vergangener aufgewendeter Selbstkontrolle ganz klar künstliche Diskretisierungen.

Analytische Sätze: Es ist nicht einfach zu klären, welche Annahmen den Status von begriffsanalytischen Notwendigkeiten haben. Ist die Annahme, dass die

Motivation durch Selbstrelevanz erhöht wird, ein analytischer Satz, der sich aus der Definition von Motivation ergibt und mithin a priori wahr ist? Oder sollte das doch empirisch überprüft werden? Ein weiterer Kandidat für diese Art von Problem ist die Frage, ob die Aufwendung von Willenskraft gleichbedeutend mit Typ-2-Verarbeitung ist.

Wie bereits erwähnt, betreffen die meisten dieser Probleme auch Theorien, die herkömmlich verbal oder grafisch formuliert sind, nur dass sie dort weniger auffallen. Es ist ein wichtiger Vorteil der formalen (Re-)Konstruktion von Theorien, dass solche Fragen nicht einfach ignoriert werden können. Inwieweit Ad-hoc-Annahmen, die in solchen Fällen oft getroffen werden, eine Bereicherung der psychologischen Theorienlandschaft darstellen, kann diskutiert werden.⁴

Schlussfolgerungen

Selbst wenn es nicht überzeugend gelungen sein mag, eine Theorie vollständig als finiten Automaten abzubilden, kann allein der Versuch als fruchtbar angesehen werden. Was sind die Gründe dafür?

Der Versuch regt dazu an, systematischer über die Zustände nachzudenken, in denen sich ein psychisches System befinden kann. Mit wie vielen Dimensionen sind diese Zustände beschreibbar? Müssen die Dimensionen unabhängig voneinander sein? Da Zustände in vielen Theorien eine Rolle spielen, kann man sie untereinander vergleichen. Wie ähnlich sind sich etwa der Zustand, in dem vorwiegend Verarbeitung vom Typ 2 stattfindet, und der Zustand, in dem das psychische System starke Selbstkontrolle ausübt? Oder: Wie ähnlich ist der Flow-Zustand dem des intrinsisch motivierten Handelns? Solche Überlegungen enthalten viel Potenzial zur Vereinigung von Theorien, indem sie an den Stellen verbunden werden, an denen man äquivalente Zustände identifiziert hat.

⁴ Als klassisches Beispiel für eine Theorie mit Ad-hoc-Annahmen kann die dynamische Handlungstheorie von Atkinson und Birch (1970) angeführt werden. Aber auch Dörners PSI-Theorie (Dörner, 1998) enthält viele solcher Annahmen.

Die Auseinandersetzung mit den relevanten Dimensionen zur Zustandsbeschreibung macht deutlich, dass in psychologischen Theorien häufig nur sehr spezielle Zustände thematisiert werden. In der Willenskraft-Theorie werden zum Beispiel die Zustände sich regenerierender Ego Depletion vernachlässigt. Was ist mit solchen Zuständen? Bei Annahme unabhängiger Dimensionen gibt es viele mögliche Zustände. Sind die meisten davon uninteressant oder trivial, oder verbirgt sich dort ein Potenzial für neue Entdeckungen? Wenn man dagegen abhängige Dimensionen annimmt, ergeben sich interessante Fragen nach den Ursachen und Hintergründen der Abhängigkeit. Gibt es z. B. physiologische Mechanismen und Parameter, die charakteristisch für bestimmte Zustände sind? Auch die oben angesprochene Frage nach dem ontologischen Status von Zuständen muss nicht nur als Problem betrachtet werden, sondern auch als Chance, solche Probleme systematischer zu untersuchen.

Die beschriebene Selektivität von Zuständen betrifft in vielleicht noch größerem Ausmaß die Übergangsregeln. Es hat sich gezeigt, dass die meisten sprachlich formulierten Regeln global sind, das heißt, sie beziehen sich auf nur eine Zustandsdimension und damit immer auf eine Menge von unterscheidbaren Zuständen. Diese Erkenntnis kann eine systematischere Inspektion einzelner Zustände anregen und damit auch neue empirische Forschung.

Ganz offensichtlich ist die automatentheoretische Formalisierung psychologischer Theorien sehr mechanistisch und wenig kompatibel mit den Gepflogenheiten weiter Teile der Psychologie. Das zeigt sich in den Äußerungen von ChatGPT, das ich Anfang 2023 zu meiner Idee befragt habe:

It is possible to formalize psychological theories as ... finite automata, but it would likely be a simplification of the theory and may not capture all of the nuances and complexity of human behavior. ... These theories are typically based on observational and experimental data, and they often involve complex and nuanced concepts such as motivation, emotion, and cognition. (ChatGPT 3.5, Februar 2023)

Und in mancher Hinsicht stimme ich der Einschätzung von ChatGPT sogar zu. Dennoch hat die vorgestellte Analyse gezeigt, dass die angesprochene Nuancierung und Komplexität sich meist auf eine sehr enge Auswahl von Zuständen und deren Übergangsregeln beziehen und dass sprachliche Formulierungen gut dazu geeignet sind, Unschärfen zu verschleiern. Dabei ist der hier vorgestellte Ansatz nur eine von vielen Möglichkeiten der Modellierung. Ganz ähnliche Erfahrungen kann man bei der kognitiven Modellierung machen, etwa mit ACT-R (Anderson, 2007). Der automatentheoretische Ansatz hat allerdings den Vorteil, viel leichter erlernbar zu sein.

Letztlich stellt sich auch die Frage, ob finite Automaten ein angemessenes Modell für die menschliche Psyche sein können. Da Modelle eine reduzierte Abbildung der Wirklichkeit sind, würde ich dies bejahen – denn das bedeutet ja nicht, dass man die menschliche Psyche selbst als finiten Automaten auffasst. Man kann meines Erachtens Prozesse mechanistisch analysieren, ohne damit die Intentionalität eines handelnden Subjekts infrage zu stellen (Wendt & Funke, 2022).

Danksagung. Ich danke zwei anonymen Gutachtern sowie Carlos Kölbl für mehrere wertvolle Hinweise zur Verbesserung dieses Aufsatzes – und natürlich auch Joachim Funke für den langjährigen und anregenden wissenschaftlichen Austausch. Lieber Joachim, bleib gesund und begleite noch viele Jahre unsere Bemühungen in der Psychologie!

Referenzen

- Ach, N. (1910). *Über den Willensakt und das Temperament*. Quelle & Meyer.
- Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195324259.001.0001>
- Atkinson, J. W., & Birch, D. (1970). *The dynamics of action*. John Wiley.
- Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5), 1252–1265. <http://dx.doi.org/10.1037/0022-3514.74.5.1252>

- Baumeister, R. F., & Vohs, K. D. (2007). Self-regulation, ego depletion, and motivation. *Social and Personality Psychology Compass*, 1(1), 115–128. <https://doi.org/10.1111/j.1751-9004.2007.00001.x>
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, 16(6), 351–355. <http://dx.doi.org/10.1111/j.1467-8721.2007.00534.x>
- Borgstede, M., & Eggert, F. (2023). Squaring the circle: From latent variables to theory-based measurement. *Theory & Psychology*, 33(1), 118–137. <https://doi.org/10.1177/09593543221127985>
- Csikszentmihalyi, M., Abuhamdeh, S., & Nakamura, J. (2014). Flow. In M. Csikszentmihalyi (Hrsg.), *Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi* (S. 227–238). Springer. https://doi.org/10.1007/978-94-017-9088-8_15
- Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dörner, D. (1998). *Bauplan für eine Seele*. Rowohlt.
- Evans, J. St. B. T. (2012). Spot the difference: Distinguishing between two kinds of processing. *Mind & Society*, 11(1), 121–131. <http://dx.doi.org/10.1007/s11299-012-0104-2>
- Fahrenberg, J. (2015). *Theoretische Psychologie – Eine Systematik der Kontroversen*. Pabst Science Publishers. <https://doi.org/10.23668/psycharchives.10364>
- Frieze, M., Loschelder, D. D., Gieseler, K., Frankenbach, J., & Inzlicht, M. (2019). Is ego depletion real? An analysis of arguments. *Personality and Social Psychology Review*, 23(2), 107–131. <http://dx.doi.org/10.1177/1088868318762183>
- Gollwitzer, P. (1990). Action phases and mind-sets. In E. T. Higgins & R. M. Sorrentino (Hrsg.), *The handbook of motivation and cognition: Foundations of social behavior* (Bd. 2, pp. 53–92). Guilford Press.
- Haken, H., & Schiepek, G. (2006). *Synergetik in der Psychologie. Selbstorganisation verstehen und gestalten*. Hogrefe.
- Inzlicht, M., & Berkman, E. (2015). Six Questions for the resource model of control (and some answers). *Social and Personality Psychology Compass*, 9(10), 511–524. <https://doi.org/10.1111/spc3.12200>

- Lewin, K. (1926). Vorsatz, Wille und Bedürfnis. *Psychologische Forschung*, 7, 330–385. <https://doi.org/10.1007/BF02424365>
- McPhetres, J., Albayrak-Aydemir, N., Barbosa Mendes, A., Chow, E. C., Gonzalez-Marquez, P., Loukras, E., Maus, A., O'Mahony, A., Pomareda, C., Primbs, M. A., Sackman, S. L., Smithson, C. J. R., & Volodko, K. (2021). A decade of theory as reflected in Psychological Science (2009–2019). *PLOS ONE*, 16(3), e0247986. <https://doi.org/10.1371/journal.pone.0247986>
- Moore, E. F. (1956). Gedanken-experiments on sequential machines. In C. E. Shannon & J. MacCarthy (Hrsg.), *Automata studies* (S. 129–153). Princeton University Press. <https://doi.org/10.1515/9781400882618-006>
- Oberauer, K., & Lewandowsky, S. (2019). Addressing the theory crisis in psychology. *Psychonomic Bulletin & Review*, 26(5), 1596–1618. <https://doi.org/10.3758/s13423-019-01645-2>
- Schiepek, G., Tominschek, I., & Heinzel, S. (2014). Self-organization in psychotherapy: Testing the synergetic model of change processes. *Frontiers in Psychology*, 5(1089), 1–11. www.doi.org/10.3389/fpsyg.2014.01089
- Wendt, A. N., & Funke, J. (2022). *Wohin steuert die Psychologie?* Vandenhoeck & Ruprecht. <https://doi.org/10.13109/9783666408007>
- Wendt, A. N., & Wolfradt, U. (2022). Theoretical psychology: Discursive transformations and continuity in Psychological Research/Psychologische Forschung. *Psychological Research*, 86(8), 2321–2340. <https://doi.org/10.1007/s00426-022-01727-2>

Human Eye Movements as Indicators of Complexity

MIRIAM SPERING

University of British Columbia Vancouver, Kanada

Abstract. Perceiving and interacting with the world around us is a multimodal process that involves orienting our eyes, head, and body towards objects of interest. It is also a highly dynamic process during which the eyes continuously scan the visual environment to sample information. This active sensing allows us to overtly and rapidly react and adapt to survival-critical changes in the environment. At a more subtle and covert level, it also accompanies and supports cognitive processes that underlie planning, decision making and problem solving. This chapter discusses research advances on active sensing during cognitive tasks at varying levels of complexity and proposes that eye movements can provide direct, real-time insight into cognitive processes, at a fine spatial and temporal scale.

Our everyday life is characterized by problems that require cognitive competencies, ranging from simple decisions about what food to eat or which route to drive to complex problem solving in ill-defined and dynamic situations. Planning, deciding, and solving problems are key components of human nature, and our daily lives reflect an ongoing quest to master these competencies. A wide range of tests and tools have been developed that allow assessment of cognitive and executive functions in the relatively controlled environment of the laboratory as

well as “in the real world” (Betsch et al., 2011; Diamond, 2012) and at varying levels of complexity (Funke & Spering, 2006).

Yet, experimental research on human thinking currently does not appropriately address real-world complexity. Studies largely focus on performance outcome, and tools that allow us to investigate underlying processes and mechanisms at a granular level are less accessible. With advances in machine learning, researchers have found that functional magnetic resonance imaging can reveal and quite literally track the cognitive stages that constitute a problem solving process (Anderson et al., 2016; Anderson et al., 2020). Even though these studies employ challenging video games, these are well-defined and relatively simple tasks requiring visual skills such as time-to-contact estimation or multiple-object tracking. Similarly, studies utilizing electroencephalography during problem solving tasks have largely focused on arithmetic or insight problem solving (e.g., Bowden et al., 2005; Dietrich & Kanso, 2010; Hinault & Lemaire, 2016).

By contrast, methods for assessing the cognitive stages that underlie solving real-world problems, ill-defined by nature (e.g., Dörner & Funke, 2017), are sparse. The more complex a task, the harder it is to assess problem solving steps, strategies, and performance within the constraints of an experimental laboratory. Perhaps it is indeed time for “problem solving research [...] to give up methodological rigorism and to open the field once again for controlled introspection and single-case studies that describe the phenomena in a less restricted way” (Funke, 2014, p. 495). Even though methods such as thinking out loud might provide a rich qualitative data set, they are subjective by definition. Moreover, observers can only report thoughts that are accessible to them in any given moment. In this chapter, I will propose that human eye movements can serve a similar purpose to self-reports and sensitively reflect the strategies we develop and steps we take when we plan, decide, and solve problems. Even though research on eye movements during real-world decision making and problem solving is in its infancy, eye movements are emerging as a versatile tool to investigate cognitive processes (Ryan & Shen, 2020; Spering, 2022), including processes that occur outside of our immediate awareness (Hannula et al., 2007; Spering & Carrasco, 2015).

In the context of research on complex cognition, eye movements have at least four key advantages: (1) We move our eyes naturally and spontaneously and need little training or instruction to do so; (2) eye movements are highly sensitive, including to information that observers are not consciously aware of; (3) they are easy and relatively inexpensive to measure non-invasively; and perhaps most importantly, (4) eye movements are continuous and provide fine-grained spatial and temporal information.

Eye Movements as Metrics of Cognition: Prerequisites

When we view the world around us, our eyes move continuously. We employ an arsenal of different types of eye movements to fixate gaze on objects of interest, redirect gaze during activities such as reading or scene viewing, track objects that move, or compensate for image motion resulting from movements of our head and body. Eye movements across the vertebrate lineage are finely tuned to the image statistics of our natural environment and the demands of the tasks we face (Land & Nilsson, 2013; Land & Tatler, 2009). Humans make three to five eye movements per second, making it the most common movement primates engage in (Bargary et al., 2017). Yet, the extraocular muscles that move the eyes do not fatigue (Fuchs & Binder, 1983), allowing us to move our eyes continuously. It is this continuity at high accuracy and precision that makes eye movements such an appealing model system for the inference of cognitive processes. Humans use a combination of saccades (rapid gaze shifts to objects of interest), smooth pursuit (slow, continuous tracking of moving objects), vergence (movements to focus in depth/3D), and movements that compensate for self-motion, such as the vestibulo-ocular reflex (Leigh & Zee, 2015). These larger eye movements are interspersed with fixation, periods of relative rest, during which the eyes move within a small spatial range (Poletti & Rucci, 2016; Rolfs, 2009). Eye movements align the fovea (the area of highest photoreceptor density and, thus, highest visual acuity) with objects of interest and allow us to see the visual world in all its richness and spatial detail (Gegenfurtner, 2016). They support and guide isolated motor actions such as

reaching or walking (de Brouwer et al., 2021; Matthis et al., 2018) and accompany everyday tasks ranging from making a sandwich to driving a car (Foulsham, 2015; Hayhoe, 2017; Hayhoe & Ballard, 2005; Lappi, 2016); they are also believed to contribute to high performance in sports (Vickers, 2016).

The discovery that cognitive processes, and specifically task instructions, influence eye movements was first reported by Russian vision scientist Alfred Yarbus (Tatler et al., 2010; Yarbus, 1967; see also Castelano et al., 2009). Yarbus made several key observations: When viewing a scene, the eyes' initial fixation is preferentially directed at persons within the scene, more specifically, at their faces (Fletcher-Watson et al., 2008) and particularly the eye region (Kingstone, 2009). During longer viewing, gaze cyclically and repeatedly revisits the most important elements of a picture. In more recent years, eye movement responses have been formalized to serve investigating cognitive processes during visual search, reading, and scene viewing (Rayner, 2009). The utility of eye movements as sensitive indicators of cognitive processes relies on the following key assumptions: (1) Observers look at objects they are processing (Rayner & Liversedge, 2004), indicating that fixation locations allow us to deduce what sort of information about an object or a location is currently being processed (eye-mind hypothesis; Just & Carpenter, 1980); (2) fixation duration reflects processing time (Kaller et al., 2009; Vickers, 2016) such that the longer an object or a location is fixated, the more information is gathered and processed; (3) in tasks that involve object manipulation and dynamic interactions with the visual environment, eye movements precede actions (de Brouwer et al., 2021), thereby reflecting advanced planning of future task-related actions; (4) eye movements not only reflect cognitive states but contribute to them as well (Ryan & Shen, 2020; Sperring, 2022).

Studies that consider eye movements as cognitive indicators typically rely on standard fixation and saccade metrics. The large majority of studies in this field uses fixation or dwell time, saccade endpoints, and the direction of saccades (e.g., forward, backward, alternating). Some studies report saccade kinematics, such as saccade latency (reaction time), peak velocity, amplitude, duration, or their vigor (a metric that calculates velocity as a function of amplitude); however, these

metrics are still underutilized. Dynamic aspects such as smooth pursuit, vergence, or the vestibulo-ocular reflex are almost never reported. This is most likely due to the nature of the stimulus material, which is largely static and presented in 2D; moreover, observers are almost always seated in front of a computer with their head stabilized. Pupil dilation and blink rate are two relatively novel markers of cognitive task features and observer characteristics; they have the potential to shed light on longer-scale or sequential cognitive processes (Ebitz & Moore, 2019; Jongkees & Colzato, 2016; for a review, see Spering, 2022). In the following sections, I will discuss selected, representative studies showcasing the utility of eye movements in tasks that measure humans' cognitive competencies at different levels of complexity.

Eye Movements Reflect Humans' Adaptive Capability to Plan

In its broadest form, planning involves an intention or decision concerning what one is going to do. This process is comprised of a series of internal goals that are formulated and stored in memory and then retrieved to guide a behavior or achieve an action and compare it against an internal model of the environment. Eye movements can reveal planning at a very low processing level, i.e., they reflect steps of sensory information accrual (Gottlieb & Oudeyer, 2018). This step includes searching for and selecting an object as a target, weighting and integrating sensory information, and preparing for and executing an action. Congruently, much of the research in this area has been conducted with highly simplified tasks that break down the cognitive planning process to its lowest level: planning on or deciding where to look next. Notwithstanding the simplicity in their design, the results of these studies have important implications for studying more complex forms of planning as well. They reveal that eye movements tend to be directed at the most salient object (the one that stands out the most) in a visual scene (Fecteau & Munoz, 2006) and depend on statistical regularities as well as the expected reward associated with a target or location (for a review see Eckstein, 2011). Eye movements during natural scene viewing typically follow a stereotypical pattern of

making a saccade to a novel object (exploration) and revisiting a previously fixated object (exploitation; Mirpour & Bisley, 2021). In this context, eye movements have often been shown to behave near-optimally, i.e., they scan a visual scene efficiently, taking into account natural scene characteristics as well as constraints of the visual system itself (Najemnik & Geisler, 2005). Moreover, the finding that eye movements can be made in anticipation of a visual event—before a target starts to move (Kowler, 2011), before it reaches a certain location in space (i.e., “look-ahead fixations”, Mennie et al., 2007), or in prediction of a future action (Flanagan & Johansson, 2003; Fookien & Spering, 2020)—implies that eye movements play an important role in motor task planning. At a somewhat more complex level, during visual-guided navigation, observers rely on an internal map (or plan) of their surroundings. For example, when navigating a virtual maze, human eye movements are rapidly deployed to sequentially review the future path, reflecting internal plans in real time (Zhu et al., 2022).

Understanding information accrual as a building block of human planning is especially important in situations that are marked by uncertainty and that change dynamically over time. After all, we cannot plan, decide, or solve problems without knowing what information is available first. Eye movements reveal the nature of the information that is being gathered. They also reflect memory representations of object or scene characteristics (Ryan & Shen, 2020), and analyzing them over time can shed light on the relative influence of and balance between low-level visual and higher-level cognitive influences (Schütt et al., 2019).

Equipped with knowledge of these building blocks, we can move on to planning tasks that require executive functions. The Tower of Hanoi (or, in its variation, the Tower of London) is a goal-stacking task that requires transferring a number of balls or disks of different colors or sizes from one peg to another following stacking-order and movement rules. It constitutes a simple, static environment that requires a mental search through a space of possible solution paths. Assessing eye movements in this task and in similar scenarios can reveal optimality in the steps taken and overall task performance. For example, the analysis of dwell time on goal-space and work-space locations in this task reveals discrete planning phases

(Hodgson et al., 2000). Eye movements are initially concentrated on the goal space, reflecting the gathering and assessment of information to create an internal representation of the goal state. During a subsequent phase of solution elaboration, eye movements focus on the work space. Errors during this phase are associated with an overall lack of eye movements towards problem-critical locations (e.g., balls that had to be moved before others could be accessed). Similarly, Ayala and colleagues (2022) showed that sub-optimal eye movement patterns in early planning stages reflect poor information extraction from the task environment. Eye movements can also reflect the use of strategies, algorithms, and routines (Patsenko & Altmann, 2010) and indicate task demands (Nitschke et al., 2012), learning (Asato et al., 2006), and memory representations or dysfunctions (Droll & Hayhoe, 2007; Huddy et al., 2007). An exploration of eye movements (fixation duration and saccade amplitudes) during the Rubik's cube task linked switches from ambient to more focal viewing to different states of attentional processing (Guo et al., 2022).

To translate these findings from active engagement with artificial tasks to active sensing in the real world, researchers could capitalize on the emergence of mobile eye-tracking methods to characterize eye movements during ecologically valid scenarios (Ladouce et al., 2017). Real-world task planning, for example, includes time pressure, cost, risk, and social factors, and whether eye movements can reflect performance and performance-related factors in these complex, dynamic social situations remains to be investigated.

Eye Movements as a Window Into Decision Making

Decisions are ubiquitous, and just like for planning and problem solving, they can be simple and binary—whether to drive through a changing traffic light or whether to stop—only requiring selecting, discriminating, and weighing the accumulated sensory evidence, or much more complex. Value-based decisions—for example, which lunch option to choose, gym to join, or stock option to invest in—require identifying and deliberating expected costs and benefits. Eye movements have

been used in both simple and more complex decision scenarios. They can indicate target selection and deliberation processes as well as performance monitoring (confidence; Ebitz & Moore, 2019). They can serve as an index of stimulus and task features such as uncertainty or value (Shadmehr et al., 2019). They signify and predict decision timing and temporal expectation (Seideman et al., 2018). They also indicate levels of cognitive control, for example, when a decision scenario involves inhibiting a response (Munoz & Everling, 2004). Notably, some of these more process-related outcome variables are derived not only based on where we look and for how long (i.e., endpoints and dwell time) but also on metrics such as saccade peak velocity. For example, the velocity of a saccade reflects decision accuracy as well as the confidence associated with that decision (Seideman et al., 2018). Confidence or decision certainty are also reflected in the pupil response, a nuanced measure that also indicates how beliefs about one's own decision change over time (e.g., Colizoli et al., 2018). The vigor of a saccadic eye movement, i.e., the time the movement takes relative to its amplitude, describes the strength, effort, or energy expenditure of a saccade. It has been related to the economic utility of a decision outcome, i.e., the interaction of the subjective reward value associated with a decision outcome, and the effort that has to be exerted to obtain this reward (Shadmehr et al., 2019). For example, when an observer deliberates between two choice alternatives that differ in value—a smaller monetary reward paid out immediately, or a larger reward paid out later—the eyes initially move with the same vigor to both options. As the decision process unfolds, vigor increases for the response alternative that is ultimately chosen. These results highlight the incredible potential of eye movements as markers of decision timing and of humans' ability to keep track of reward and effort over time. Similarly, the analysis of smooth pursuit eye movements harnesses the continuous nature of these movements. In a simulated decision task mimicking the real-world timing and requirements of baseball, Fookien and Spering (2020) found that the accuracy of pursuit and timing of saccades predicted how well observers ultimately did in this task, highlighting that eye movements not only reflect the decision outcome but potentially contribute to it as well.

Even though these studies demonstrate the impressive potential of eye movements as markers of decision processes, they all rely on relatively simple tasks. More recently, eye movements have been measured in multi-alternative, multi-attribute choice tasks that mimic requirements of selecting an item from a vending machine (Thomas et al., 2021). Findings obtained with simple tasks also appear to apply to tasks that involve moral decisions (Pärnamets et al., 2015). It is now important to apply these findings and principles to decision tasks embedded in contexts that change dynamically over time (Barendregt et al., 2022), reflecting real-world decisions that require adaptive behaviour.

Can eye Movements Reflect the Complexity of a Problem?

Problem solving, by definition, deals with situations in which steps toward the solution of the problem are not clear; in fact, the problem itself can be ill-defined. In its simplest form, a problem could be a riddle or a puzzle. In its more complex form, a problem lacks transparency, is polytelic (i.e., it might have to be approached from multiple, sometimes competing, perspectives to satisfy multiple subgoals), contains interacting components, and changes dynamically within an everchanging context (Dörner & Funke, 2017). In short: most global challenges (poverty, climate change, pandemics, racism, etc.) constitute complex problems.

It is easy to conceive that eye movements can measure simple problem solving. In fact, the first systematic studies on eye movements during cognitive tasks in general compared expert and novice chess players—with chess considered the “*drosophila*” of cognitive psychology—and identified distinguishable eye movement strategies that marked the success of experts (Chase & Simon, 1973; de Groot, 1978). Experts’ eye movements tended to alternate more often between key pieces and fixate less on individual pieces, indicating that chess masters encode visual information in chunks rather than in individual units (Reingold et al., 2001). A larger visual span allows not only retaining more moves in visual short-term memory but also planning more moves ahead. These findings show that eye movements reflect the perceptual organization and internal representation

of task-relevant information in situations that go beyond highly simplified visual search tasks. Similarly, eye movements reflect the acquisition of knowledge toward a solution (sometimes termed the “Aha” moment) in anagram problem solving tasks; a pattern of decreased fixations on distractor letters can even predict the point where the observer is ready to solve the task (Ellis et al., 2011). Recent studies have shown that the “Aha” moment is often preceded by a reduction in visual information accrual, evidenced by more frequent eye blinks and fewer eye movements into the problem space (Salvi et al., 2015).

Yet, none of these studies involve complex tasks or even dynamic changes to the solution space. Can eye movements indicate the complexity of a task, and of the thought process of a problem solver? The studies summarized in this chapter show that eye movements can indeed reveal the efficiency with which an observer gathers visual information across different levels of complexity, from binary and highly simplified search tasks to simple problems. Under uncertainty or lack of task transparency—one of the hallmarks of a complex problem—information gathering and integration becomes even more critical, and eye movements can be a valuable tool to help us understand sensory information accrual when information is sparse. Eye movements, especially pupil responses, also reflect task difficulty and the cognitive demands a task poses—characteristics that can be expected to become more pronounced with increasing task complexity. Eye movements could be analyzed in tasks that comprise multiple simultaneous goals, perhaps with a focus on which goals observers prioritize or identify as most task relevant.

To design tasks that allow investigating eye movements during complex problem solving we can look to research on human factors—of how humans interact with elements of their workplace or technical environment. Even though these tasks are not always complex, they are marked by features that contribute to task complexity, such as dynamics, uncertainty, urgency, and high risk. Eye movements have been analyzed in medical contexts, such as in surgeons in the operating room or radiologists during diagnostic pathology (e.g., Pauszek, 2023) or in the automotive industry and in drivers during multitasking (Ahlström et al., 2021). Even though the focus of this work has overwhelmingly been on person factors such

as workload, vigilance, attention, and fatigue, or on object factors such as device usability and equipment design, these studies show that eye movements can reveal processes and outcomes in real-world dynamic social contexts. Moreover, virtual reality displays provide an opportunity to create immersive and ecologically valid scenarios and track eye, hand, and body movements simultaneously (Anderson et al., 2023).

In summary: even though eye movement studies do not yet capture and reflect the complexity of real-world problem solving scenarios, they provide insights into the fundamental building blocks of human cognition. Since the first systematic eye tracking studies on human cognition, our knowledge of the visual signals that guide human eye movements has increased dramatically. Cognitive psychologists can now harness this knowledge and disentangle the low-level visual drivers that draw our eyes to salient objects in a scene from the higher-level cognitive factors that determine strategies and processes underlying complex cognition.

References

- Ahlström, C., Kircher, K., Nyström, M., & Wolfe, B. (2021). Eye tracking in driver attention research—How gaze data interpretations influence what we learn. *Frontiers in Neuroeconomics*, 2. <https://doi.org/10.3389/fnrgo.2021.778043>
- Anderson, J. R., Betts, S., Fincham, J. M., Hope, R., & Walsh, M. W. (2020). Reconstructing fine-grained cognition from brain activity. *Neuroimage*, 221, 116999. <https://doi.org/10.1016/j.neuroimage.2020.116999>
- Anderson, J. R., Pyke, A. A., & Fincham, J. M. (2016). Hidden stages of cognition revealed in patterns of brain activation. *Psychological Sciences*, 27, 1215–1226. <https://doi.org/10.1177/0956797616654912>
- Anderson, N. C., Bischof, W. F., & Kingstone, A. (2023). Eye tracking in virtual reality. *Current topics in behavioral neurosciences*. Heidelberg: Springer. https://doi.org/10.1007/7854_2022_409
- Asato, M. R., Sweeney, J. A., & Luna, B. (2006). Cognitive processes in the development of TOL performance. *Neuropsychologia*, 44, 2259–2269. <https://doi.org/10.1016/j.neuropsychologia.2006.05.010>

- Ayala, N., Zafar, A., & Niechwiej-Szwedo, E. (2022). Gaze behaviour: A window into distinct cognitive processes revealed by the Tower of London test. *Vision Research*, 199, 108072. <https://doi.org/10.1016/j.visres.2022.108072>
- Barendregt, N. W., Gold, J. I., Josić, K., & Kilpatrick, Z. P. (2022). Normative decision rules in changing environments. *eLife*, 11, e79824. <https://doi.org/10.7554/eLife.79824>
- Betsch, T., Funke, J., & Plessner, H. (2011). Problemlösen: Paradigmen und Methoden. In T. Betsch, J. Funke, & H. Plessner (Eds.), *Denken – Urteilen, Entscheiden, Problemlösen* (pp. 191–199). Springer. https://doi.org/10.1007/978-3-642-12474-7_15
- Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Science*, 9, 322–328. <https://doi.org/10.1016/j.tics.2005.05.012>
- Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3), Article 6, 1–15. <https://doi.org/10.1167/9.3.6>
- Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual information processing* (pp. 215–281). Academic Press. <https://doi.org/10.1016/B978-0-12-170150-5.50011-1>
- Colizoli, O., De Gee, J. W., Urai, A. E., & Donner, T. H. (2018). Task-evoked pupil responses reflect internal belief states. *Scientific Reports*, 8, 1–13. <https://doi.org/10.1038/s41598-018-31985-3>
- De Brouwer, A. J., Flanagan, J. R., & Spering, M. (2021). Functional use of eye movements for an acting system. *Trends in Cognitive Sciences*, 25, 252–263. <https://doi.org/10.1016/j.tics.2020.12.006>
- De Groot, A. D. (1978). *Thought and choice in chess* (2nd ed.). Mouton.
- Diamond, A. (2012). Executive functions. *Annual Review of Psychology*, 64, 135–168. <https://doi.org/10.1146/annurev-psych-113011-143750>
- Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological Bulletin*, 136, 822–848. <https://doi.org/10.1037/a0019749>
- Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8(1153), 1–11. <https://doi.org/10.3389/fpsyg.2017.01153>

- Droll, J. A., & Hayhoe, M. M. (2007). Trade-offs between gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 1352–1365. <https://doi.org/10.1037/0096-1523.33.6.1352>
- Ebitz, R. B., & Moore, T. (2019). Both a gauge and a filter: Cognitive modulations of pupil size. *Frontiers in Neurology*, 9, 1190. <https://doi.org/10.3389/fneur.2018.01190>
- Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, 11(5), 14. <https://doi.org/10.1167/11.5.14>
- Ellis, J. J., Glaholt, M. G., & Reingold, E. M. (2011). Eye movements reveal solution knowledge prior to insight. *Consciousness and Cognition*, 20, 768776. <https://doi.org/10.1016/j.concog.2010.12.007>
- Fecteau, J. H., & Munoz, D. P. (2006). Saliency, relevance, and firing: A priority map for target selection. *Trends in Cognitive Science*, 10, 382–390. <https://doi.org/10.1016/j.tics.2006.06.011>
- Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception*, 37, 571–583. <https://doi.org/10.1068/p5705>
- Fooken, J., & Spering, M. (2020). Eye movements as a readout of sensorimotor decision processes. *Journal of Neurophysiology*, 123, 1439–1447. <https://doi.org/10.1152/jn.00622.2019>
- Foulsham, T. (2015). Eye movements and their functions in everyday tasks. *Eye*, 29, 196–199. <https://doi.org/10.1038/eye.2014.275>
- Funke, J. (2014). Problem solving: What are the important questions? In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 493–498). Cognitive Science Society.
- Funke, J., & Spering, M. (2006). Methoden der Denk-und Problemlöseforschung. In J. Funke (Ed.), *Denken und Problemlösen* (pp. 647–744). Hogrefe.
- Gegenfurtner, K. R. (2016). The interaction between vision and eye movements. *Perception*, 45, 1333–1357. <https://doi.org/10.1177/0301006616657097>
- Gottlieb, J., & Oudeyer, P.-Y. (2018). Toward a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, 19, 758–770. <https://doi.org/10.1038/s41583-018-0078-0>

- Guo, Y., Helmert, J. R., Graupner, S.-T., & Pannasch, S. (2022). Eye movement patterns in complex tasks: Characteristics of ambient and focal processing. *PLoS One*, *17*(11), e0277099. <https://doi.org/10.1371/journal.pone.0277099>
- Hannula, D. E., Ryan, J. D., Tranel, D., & Cohen, N. J. (2007). Rapid onset relational memory effects are evident in eye movement behavior, but not in hippocampal amnesia. *Journal of Cognitive Neuroscience*, *19*, 1690–1705. <https://doi.org/10.1162/jocn.2007.19.10.1690>
- Hayhoe, M. M. (2017). Vision and action. *Annual Review of Vision Science*, *3*, 389–413. <https://doi.org/10.1146/annurev-vision-102016-061437>
- Hayhoe, M., & Ballard, D. (2005). Eye movements in natural behavior. *Trends in Cognitive Science*, *9*, 188–194. <https://doi.org/10.1016/j.tics.2005.02.009>
- Hinault, T., & Lemaire, P. (2016). What does EEG tell us about arithmetic strategies? A review. *International Journal of Psychophysiology*, *106*, 115–126. <https://doi.org/10.1016/j.ijpsycho.2016.05.006>
- Hodgson, T. L., Bajwa, A., Owen, A. M., & Kennard, C. (2000). The strategic control of gaze direction in the Tower of London task. *Journal of Cognitive Neuroscience*, *12*, 894–907. <https://doi.org/10.1162/089892900562499>
- Huddy, V. C., Hodgson, T. L., Harrison, I., Kapasi, M., Stanley, H., Thomas, M., Barnes, T. R. E., & Joyce, E. M. (2007). Gaze strategies during planning in first-episode psychosis. *Journal of Abnormal Psychology*, *116*, 589–598. <https://doi.org/10.1037/0021-843X.116.3.589>
- Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*, 441–480. [https://doi.org/10.1016/0010-0285\(76\)90015-3](https://doi.org/10.1016/0010-0285(76)90015-3)
- Kaller, C. P., Rahm, B., Bolkenius, K., & Unterrainer, J. M. (2009). Eye movements and visuospatial problem solving: Identifying separable phases of complex cognition. *Psychophysiology*, *46*, 818–830. <https://doi.org/10.1111/j.1469-8986.2009.00821.x>
- Kingstone, A. (2009). Taking a real look at social attention. *Current Opinion in Neurobiology*, *19*, 52–56. <https://doi.org/10.1016/j.conb.2009.05.004>
- Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, *51*, 1457–1483. <https://doi.org/10.1016/j.visres.2010.12.014>
- Ladouce, S., Donaldson, D. I., Dudchenko, P. A., & Ietswaart, M. (2017). Understanding minds in real-world environments: Toward a mobile cognition approach. *Frontiers in Human Neuroscience*, *10*, 694. <https://doi.org/10.3389/fnhum.2016.00694>

- Land, M. F., & Nilsson, D.-E. (2013). *Animal eyes*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199581139.001.0001>
- Land, M. F., & Tatler, B. W. (2009). *Looking and acting: Vision and eye movements in natural behaviour*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198570943.001.0001>
- Lappi, O. (2016). Eye movements in the wild: Oculomotor control, gaze behavior & frames of reference. *Neuroscience & Biobehavioral Reviews*, 69, 49–68. <https://doi.org/10.1016/j.neubiorev.2016.06.006>
- Leigh, R. J., & Zee, D. S. (2015). *The neurology of eye movements*. Oxford University Press. <https://doi.org/10.1093/med/9780199969289.001.0001>
- Matthis, J. S., Yates, J. L., & Hayhoe, M. M. (2018). Gaze and the control of foot placement when walking in natural terrain. *Current Biology*, 28, 1224–1233. <https://doi.org/10.1016/j.cub.2018.03.008>
- Mennie, N., Hayhoe, M., & Sullivan, B. (2007). Look-ahead fixations: Anticipatory eye movements in natural tasks. *Experimental Brain Research*, 179, 427–442. <https://doi.org/10.1007/s00221-006-0804-0>
- Mirpour, K., & Bisley, J. W. (2021). The roles of the lateral intraparietal area and frontal eye field in guiding eye movements in free viewing search behavior. *Journal of Neurophysiology*, 125, 2144–2157. <https://doi.org/10.1152/jn.00559.2020>
- Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, 5, 218–228. <https://doi.org/10.1038/nrn1345>
- Nitschke, K., Ruh, N., Kappler, S., Stahl, C., & Kaller, C. P. (2012). Dissociable stages of problem solving (I): Temporal characteristics revealed by eye-movement analyses. *Brain and Cognition*, 80, 160–169. <https://doi.org/10.1016/j.bandc.2012.05.003>
- Pärnamets, P., Johansson, P., Hall, L., Balkenius, C., Spivey, M. J., & Richardson, D. C. (2015). Biasing moral decisions by exploiting the dynamics of eye gaze. *Proceedings of the National Academy of Sciences*, 112, 4170–4175. <https://doi.org/10.1073/pnas.1415250112>
- Patsenko, E. G., & Altmann, E. M. (2010). How planful is routine behavior? A selective-attention model of performance in the Tower of Hanoi. *Journal of Experimental Psychology: General*, 139, 95–116. <https://doi.org/10.1037/a0018268>

- Pauszek, J. R. (2023). An introduction to eye tracking in human factors healthcare research and medical device testing. *Human Factors in Healthcare*, 3, 100031. <https://doi.org/10.1016/j.hfh.2022.100031>
- Poletti, M., & Rucci, M. (2016). A compact field guide to the study of microsaccades: Challenges and function. *Vision Research*, 118, 83–97. <https://doi.org/10.1016/j.visres.2015.01.018>
- Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457–1506. <https://doi.org/10.1080/17470210902816461>
- Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science*, 12, 48–55. <https://doi.org/10.1111/1467-9280.00309>
- Rolfs, M. (2009). Microsaccades: Small steps on a long way. *Vision Research*, 49, 2415–2441. <https://doi.org/10.1016/j.visres.2009.08.010>
- Ryan, J. D., & Shen, K. (2020). The eyes are a window into memory. *Current Opinion in Behavioral Sciences*, 32, 1–6. <https://doi.org/10.1016/j.cobeha.2019.12.014>
- Salvi, C., Bricolo, E., Franconeri, S. L., Kounios, J., & Beeman, M. (2015). Sudden insight is associated with shutting out visual inputs. *Psychonomic Bulletin & Review*, 22, 1814–1819. <https://doi.org/10.3758/s13423-015-0845-0>
- Schütt, H. H., Rothkegel, L. O. M., Trukenbrod, H. A., Engbert, R., & Wichmann, F. A. (2019). Disentangling bottom-up versus top-down and low-level versus high-level influences on eye movements over time. *Journal of Vision*, 19(3), 1. <https://doi.org/10.1167/19.3.1>
- Shadmehr, R., Reppert, T. R., Summerside, E. M., Yoon, T., & Ahmed, A. A. (2019). Movement vigor as a reflection of subjective economic utility. *Trends in Neuroscience*, 42, 323–336. <https://doi.org/10.1016/j.tins.2019.02.003>
- Spering, M. (2022). Eye movements as a window into decision-making. *Annual Review of Vision Science*, 8, 427–448. <https://doi.org/10.1146/annurev-vision-100720-125029>
- Spering, M., & Carrasco, M. (2015). Acting without seeing: eye movements reveal visual processing without awareness. *Trends in Neurosciences*, 38, 189–258. <https://doi.org/10.1016/j.tins.2015.02.002>

- Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B.M. (2010). Yabus, eye movements, and vision. *i-Perception*, *1*, 7–27. <https://doi.org/10.1068/i0382>
- Thomas, A. W., Molter, F., & Krajbich, I. (2021). Uncovering the computational mechanisms underlying many-alternative choice. *eLife*, *10*, e57012. <https://doi.org/10.7554/eLife.57012>
- Vickers, J. N. (2016). Origins and current issues in Quiet Eye research. *Current Issues in Sport Science*, *1*. https://doi.org/10.15203/CISS_2016.101
- Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press. <https://doi.org/10.1007/978-1-4899-5379-7>
- Zhu, S., Lakshminarasimhan, N. A., & Angelaki, D. E. (2022). Eye movements reveal spatiotemporal dynamics of visually-informed planning in navigation. *eLife*, *11*, e73097. <https://doi.org/10.7554/eLife.73097>

Verzeichnis der Autorinnen und Autoren

IN IHREM VERHÄLTNIS ZU JOACHIM FUNKE

Dorothee Amelung completed her PhD on the psychology of climate change as part of Joachim Funke's working group in the Marsilius project "Global Governance of Climate Engineering", particularly working on individuals' perceptions of and reactions to the uncertainties and risks of controversial Climate Engineering technologies within the wider context of climate change.

Jens Beckmann and Joachim Funke first met over 35 years ago, mainly through their shared interest in cognition-oriented research on complex problem solving. At that time, Jens Beckmann was in the process of completing his postgraduate thesis (Diplomarbeit), while Joachim Funke was working on his habilitation. In 1991 Jens Beckmann joined Joachim Funke's research group at the University of Bonn for about two years, where he conducted research on learning processes in complex problem solving.

Christine Blech studied psychology at the University of Heidelberg from 2000 to 2006, followed by her PhD in 2011 at the Department of General and Theoretical Psychology in Heidelberg. Joachim Funke—who inspired her early on with cognitive psychology—was the supervisor of both her diploma thesis (*On the reactivity of causal diagram analyses in complex problem solving*) and her doctoral thesis (*Complex problem solving in situations with conflicting goals*).

Jürgen Bredenkamp und Joachim Funke trafen sich erstmals 1980. Trotz einer bereits zugesagten Stelle in Bern entschied sich Joachim Funke für Trier, was den Beginn einer 15-jährigen Zusammenarbeit markierte. In dieser Zeit entwickelte er seine Dissertation über komplexes Problemlösen, die Jürgen Bredenkamp besonders durch den innovativen Einsatz formaler Modelle beeindruckte. Als er 1984 nach Bonn wechselte, folgte ihm Joachim Funke, um dort seine Forschung weiter voranzutreiben. Obwohl ihre Forschungsinteressen unterschiedlich sind, teilen sie eine Leidenschaft für theoriegeleitete Experimente.

Axel Buchner war im Jahr 1981 Student an der Universität Trier, als er Joachim Funke als Dozent eines Experimentalpraktikums kennen und schätzen lernte. Danach trennten sich ihre Wege, aber nicht für lange. Von 1989 bis 1993 arbeitete Axel Buchner bei Joachim Funke als Mitarbeiter in spannenden Drittmittelprojekten an der Universität Bonn.

Benő Csapó (1953–2023) was Professor Emeritus at the University of Szeged, Hungary. He was the head of the Doctoral School of Education, University of Szeged, the Research Group on the Development of Competencies, Hungarian Academy of Sciences, and the Center for Research on Learning and Instruction. Benő Csapó and Joachim Funke first met in 2007, when Benő Csapó was invited to comment on the Schwerpunktprogramm “Kompetenzdiagnostik” as external expert. Since that time, both of them worked together in several international education assessment projects (e.g., PISA 2012, Assessment and Teaching of 21st Century Skills) and also became close personal friends, visiting each other regularly in Heidelberg and Szeged.

Amory Danek. Joachim Funke hat den wissenschaftlichen Werdegang von Amory Danek bis heute begleitet: Von einer ersten Einladung ins Heidelberger Kolloquium vor dreizehn Jahren, um ihre Doktorarbeit vorzustellen, bis zur Betreuung der kürzlich abgeschlossenen Habilitation *A comprehensive perspective on insight problem solving: Phenomenology and process* an der Universität Heidelberg. Auch wenn eine umfassende Theorie noch aussteht, die einsichtsvolles Problemlösen und komplexes Problemlösen vereint, so war doch der gegenseitige Austausch bezüglich dieser Themen enorm bereichernd. Viel Freude macht auch weiterhin die Zusammenarbeit bei der Herausgabe des *Journal of Dynamic Decision Making*. Menschlich ist Joachim Funke ihr mit seiner warmherzigen und integren Art ein großes Vorbild. Keiner kann wie er so engagiert und gleichzeitig unterhaltsam Wissenschaft betreiben und beschreiben und dabei bei den Zuhörern „den Funke(n)“ überspringen lassen!

Shira Elqayam is a Full Professor of Cognitive Science and Cognitive Psychology at De Montfort University, UK. Her research field is in reasoning, decision-making, and rationality. She is Associate Editor of *Thinking and Reasoning*, the specialist journal of the field. She has published over fifty articles and chapters, and led on grants from major funders including the Economic and Social Research Council, The Leverhulme Trust, and The Experimental Psychology Society. In recent years, her work has focused on meta-reasoning, and specifically on the meta-reasoning mechanisms underlying the Inferentialist theory of conditionals.

Edgar Erdfelder war Joachim Funkes erster Büronachbar zu Beginn seiner akademischen Laufbahn als wissenschaftlicher Angestellter an der Universität Trier im Jahr 1981. Aus dieser Büronachbarschaft erwuchs eine nunmehr mehr als 40-jährige Freundschaft. Folglich dauerte diese Nachbarschaft nicht nur in den Trierer Jahren an (1981–1984), sondern hatte – von einer kurzen Unterbrechung abgesehen – aufgrund beiderseitiger Bemühungen auch in der sich anschließenden gemeinsamen Zeit im akademischen Mittelbau des Psychologischen Instituts der Universität Bonn Bestand. Joachim Funkes Ruf auf den Lehrstuhl für Allgemeine und Theoretische Psychologie an der Universität Heidelberg sorgte 1997 dann allerdings für eine Unterbrechung, die Edgar Erdfelder erst 2002 durch Übernahme

des Lehrstuhls für Psychologie III an der Universität Mannheim zumindest zum Teil wieder beenden konnte. Die Büronachbarschaft konnte so immerhin durch eine Ortsnachbarschaft ersetzt werden, die bis in die Gegenwart regelmäßige Treffen und kontinuierlichen Austausch ermöglicht.

Klaus Fiedler is a Full Professor of Psychology, affiliated to Heidelberg University. He is a member of the Heidelberg Academy of Science and of the German National Academy of Science, Leopoldina, a member of various international science associations, and a Principal Investigator who won several science awards and almost 40 successful research proposals. Although his sphere of research is broad and varied, over the last decade his work has concentrated on judgment and decision making, anchored in a cognitive-ecological perspective on adaptive cognition. Central to this research area is his work on meta-cognitive myopia, conceived as a major obstacle in the way of rational decision making and action.

Helen Fischer completed her PhD in cognitive Psychology in Joachim Funke's working group in a project on the cognitive psychology of climate change, particularly focusing on climate change as a complex and non-linear system. She later worked in Joachim Funke's group as a postdoctoral researcher, focusing on the accuracy of metacognitive insight into individuals' understanding of climate change.

Annemarie Fritz-Stratmann ist Professorin für Pädagogische Psychologie sowie Gründerin und erste Vorsitzende der Akademie Wort + Zahl e. V. Sie und Joachim Funke kennen sich fast ein ganzes Leben. So war es ihr nach eigener Aussage eine besondere Freude, 55 Jahre später im Generationen-Trio mit Walter Hussy, dem Mentor, der bei Annemarie Fritz-Stratmann und Joachim Funke das Interesse am Problemlösen geweckt hat, und David Tobinski, dem gemeinsamen Doktoranden, der die Ansätze beider Professoren weiter vertieft hat, einen Beitrag zu dieser Festschrift zu leisten. Wissenschaftlich sind Annemarie Fritz-Stratmann und Joachim Funke unterschiedliche Wege gegangen, doch vor ca. 30 Jahren hat sie das Interesse an Planungsprozessen noch einmal zusammengebracht und daraus ist das Bändchen *Neue Konzepte und Instrumente zur Planungsdiagnostik* entstanden.

Was sie Joachim Funke mitgeben möchte: „Carpe Diem – von James Joyce wissen wir, wie lang ein Tag sein kann“.

Cleotilde Gonzalez is Full Research Professor of Cognitive Decision Science at Carnegie Mellon University. She is also the Founding Director of the Dynamic Decision Making (DDM) Laboratory and the Co-Director of the AI institute for Societal Decision Making. Joachim Funke's work on complex problem-solving came to her attention in the early stages of her career, while discovering the area of DDM. His innovations on the use of microworlds for experimentation were critical to the foundations of DDM. She invited Joachim Funke to visit Carnegie Mellon and give a talk in her department. Around the year 2004 they met in person and remained in contact ever since. The visits to Heidelberg, the campus, his office, the cafes, and the conversations and interactions with his students will stay in her mind forever!

Norbert Groeben ist Universitätsprofessor a. D. für Allgemeine und Kultur-Psychologie an der Uni Köln und Honorarprofessor der Universität Mannheim (seit 1993) und der Universität Heidelberg (seit 2012). Bekannt ist er auch unter den Pseudonymen Ben Roeg, Peta Panta und Egon Erb. Anlässlich seines Rufes nach Köln (1994) hat er eine Erweiterung der Denomination des Heidelberger Lehrstuhls erwirkt, auf die sich Joachim Funke erfolgreich beworben hat: Allgemeine und Theoretische Psychologie. Joachim Funke ist also der Nachfolger von Norbert Groeben in Heidelberg.

Susanne Guski-Leinwand. Als Hochschullehrer und Doktorvater ist Joachim Funke für Susanne Guski-Leinwand ein wichtiger Mensch in ihrem akademischen Leben: Er betreute ihre Diplomarbeit (1999) und Promotion (2007) und eröffnete damit für sie einen vorher so nicht angedachten Weg zur *venia legendi* für Psychologie und Geschichte der Psychologie (Friedrich-Schiller-Universität Jena, 2014). Joachim Funkes Unterstützungen wirkten nicht nur symbolisch oft wie ein „Funkenschlag“ bzw. „Funkenschläge“: Verbunden mit den Ergebnissen der historischen Forschung in ihrer Promotion unterstützte er sie durch die Vergabe von Lehraufträgen und mehrte für die Studierenden des Heidelberger Psychologischen

Instituts darüber auch das Wissen zur Ethik in der Psychologie mit historischen Bezügen (2007–2012).

Hilde Haider. Joachim Funke war für Hilde Haider stets ein anregender Diskussionspartner auf Tagungen und anderen wissenschaftlichen Veranstaltungen. Diese Diskussionen erlaubten es ihr insbesondere, verschiedene Gesichtspunkte im Bereich des komplexen Problemlösens in fruchtbarer Weise neu zu durchdenken und in ihre eigenen Forschungsarbeiten zu integrieren.

Daniel V. Holt wechselte nach dem Psychologiestudium an der University of Edinburgh (UK) zur Promotion an die Universität Heidelberg, wo er am Lehrstuhl für Allgemeine und Theoretische Psychologie unter der Leitung von Joachim Funke zunächst als Stipendiat und später als wissenschaftlicher Mitarbeiter tätig war. Von Anfang an begeisterten ihn die intellektuelle Vielfalt, die konstruktiv-kritische Auseinandersetzung mit Psychologie als Wissenschaft sowie der Schwerpunkt auf anwendungsnaher Grundlagenforschung, die für Joachim Funks eigene Arbeiten aber auch die Abteilung als Ganzes charakteristisch waren.

Walter Hussy. Joachim Funke kam im Verlauf des Vordiplomstudiums an die damals noch junge Universität Trier, wo er Walter Hussy sofort als interessierter und engagierter Student auffiel. Hier absolvierte Joachim Funke nicht nur den Diplom-Studiengang, sondern schloss auch erfolgreich die Promotion ab. Nach dem Vordiplom arbeitete er als studentische und danach als wissenschaftliche Hilfskraft in Hussys Abteilung mit und leitete als wissenschaftlicher Mitarbeiter sein langjähriges Forschungsprojekt „Entwicklung informationsverarbeitender Strukturen und Prozesse“, finanziert von der Stiftung Volkswagenwerk. Er nutzte die gewonnenen Erkenntnisse und Erfahrungen als Sprungbrett für seine weiteren Karriereschritte nach Bonn und Heidelberg.

Annette Kämmerer war bis 2014 am Psychologischen Institut der Universität Heidelberg als Akademische Direktorin tätig. Sie ist approbierte Psychotherapeutin (Verhaltenstherapie). Joachim Funke war ihr über alle Jahre ihrer Tätigkeit am Psychologischen Institut ein wichtiger und immer hilfsbereiter und interessier-

ter Kollege. Sein Humor, seine ausgleichende Art, sein gesellschaftspolitisches Engagement waren ihr eine große Bereicherung.

Jochen Musch hat Joachim Funke in Bonn kennengelernt und von ihm gelernt, dass Psychologie und Computer ganz wunderbar zusammenpassen – und dass Gummibärchen als Gegenstand psychologischer Forschung allzu lange sträflich vernachlässigt wurden. Als treuer Leser des berühmten Funke-Blogs freut er sich über alle Neuigkeiten aus dem Heidelberger Institut, an dem er selbst einst studiert hat. Häufig hat ihn dabei das mitreißende und vorbildhafte Engagement von Joachim Funke in allen Lebenslagen, aber auch die ermutigende und konstruktiv-kritische Haltung, die Joachim Funke kennzeichnet, inspiriert. Er ist der Meinung: Von Jofus geheimer Superkraft, gelegentlich auch einfach Lob zu äußern, wenn es verdient ist, und große und kleine Erfolge zu feiern, statt über ihr Ausbleiben zu klagen, können und sollten wir alle lernen.

Wolfgang Schoppek ist seit Beginn seiner wissenschaftlichen Tätigkeit mit dem Gebiet des komplexen Problemlösens vertraut: Er studierte bei Dietrich Dörner, war an dessen Bamberger Lehrstuhl mehrere Jahre als studentische Hilfskraft tätig und promovierte bei Wiebke Putz-Osterloh. Der erste persönliche Kontakt mit Joachim Funke fand in den frühen 1990er-Jahren auf einem Symposium statt, zu dem Dörner in die MP-Projektgruppe „Kognitive Anthropologie“ nach Berlin eingeladen hatte. Der Kontakt intensivierte sich 2010, als Funke Schoppek zum Workshop „Scientific computing for the cognitive sciences“ einlud und gipfelt gegenwärtig in der kontinuierlichen Zusammenarbeit bei der Herausgabe des *Journal of Dynamic Decision Making*, dessen Herausgeberteam Schoppek seit 2021 leitet. Nach Schoppeks Dafürhalten hat Joachim Funke die Forschung zum komplexen Problemlösen vor einem in den 2000er-Jahren drohenden Aus bewahrt, indem er immer wieder für kreative neue Impulse gesorgt hat.

Miriam Spring is Associate Professor and Associate Dean for Graduate and Postdoctoral Education in the Faculty of Medicine at the University of British Columbia in Vancouver, Canada. She received her diploma degree in psychology at the Ruprecht-Karls University Heidelberg, where she completed her final thesis on the role of emotions in complex problem solving under JoFu's supervision in

2002. She continued working with JoFu as a research assistant for another year and credits JoFu with giving her the freedom to explore topics at the intersection of perception and cognition and for teaching her how to ask interesting scientific questions. Even though she moved away from cognitive psychology during her PhD, she sees a lot of overlap between her current work and JoFu's. Her chapter in this volume is an attempt to build a bridge between her and JoFu's work, and is dedicated to JoFu with gratitude.

Matthias Stadler, der am LMU Klinikum in München als Professor für Learning Analytics in Medicine tätig ist, verbindet eine tiefe fachliche Wertschätzung und Inspiration mit Joachim Funke. Seitdem Matthias Stadler während seines Studiums auf die Arbeiten von Funke stieß, war er von dessen innovativen Ansätzen zur Lösung komplexer Probleme begeistert. Diese Faszination hat nicht nur sein eigenes Forschungsinteresse geprägt, sondern auch seine akademische Laufbahn maßgeblich beeinflusst. Durch die kontinuierliche Verfolgung von Funkes Arbeiten und die Auseinandersetzung mit den psychometrischen Ansätzen im Bereich des komplexen Problemlösens hat Stadler wichtige Impulse für seine eigene Forschung gewonnen. Die Möglichkeit, zu Funkes Festschrift beizutragen, ehrt ihn daher besonders, da es ihm erlaubt, seine Anerkennung und den Einfluss von Funkes Werk auf seine professionelle Entwicklung zu unterstreichen.

Robert J. Sternberg has known Joachim Funke for many years. Sternberg first became aware of Funke's work through their mutual connection to Peter Frensch, formerly Sternberg's graduate student and more recently President of Humboldt University. Sternberg and Frensch edited a book on complex problem solving, Frensch and Funke edited a book on the European approach to complex problem solving, so it was only natural that Sternberg and Funke would later team up to edit a textbook on the psychology of human thought. Funke was the scholar who nominated Sternberg for his honorary professorship at Heidelberg, and they have been great friends for many years. Sternberg always made sure to connect with Funke when he visited Germany, which has been frequently, because Sternberg's wife Karin is German and a double graduate of Heidelberg University in psychology. Sternberg greatly admires Funke not only for his ground-breaking

work on complex problem solving, but for being one of the finest human beings he has ever had the pleasure to know.

Lisa von Stockhausen lernte Joachim Funke in ihrem Psychologiestudium an der Universität Bonn kennen, als sie als studentische Hilfskraft am Lehrstuhl von Jürgen Bredenkamp arbeitete. Diese Tätigkeit und die Zusammenarbeit mit Joachim Funke (und gleich nebenan Edgar Erdfelder) weckten ihr Interesse an eigener wissenschaftlicher Forschung. Nachdem Joachim Funke 1997 den Lehrstuhl für Allgemeine und Theoretische Psychologie an der Universität Heidelberg übernommen hatte, wechselte sie als wissenschaftliche Mitarbeiterin von der Universität Trier zu Joachim Funke nach Heidelberg und promovierte und habilitierte unter seiner Betreuung. Die Begeisterung für Wissenschaft und Erkenntnis, die Joachim Funke verkörpert, hat sie zutiefst geprägt und inspiriert. Für die Freiheit, an seinem Lehrstuhl eigene Forschungsbereiche entwickeln und verfolgen zu können, ist sie ihm bis heute sehr dankbar!

David Tobinski wurde von Annemarie Fritz-Stratmann und Walter Hussy promoviert. Sehr früh kam es hierbei zur ersten persönlichen Begegnung mit Joachim Funke in Heidelberg. Neben einer geteilten Begeisterung über Fragen zur Komplexität und zum Informationsbegriff, kommt von David Tobinski der tiefe Respekt vor Joachim Funkes Ausfüllung einer Professur zur theoretischen Psychologie hinzu. Aktuell stehen Joachim Funke und David Tobinski vor dem wohl komplexesten aller Probleme: „Frieden schaffen ohne Waffen“ (Joachim Funke Blog 21.02.2025).

Alexander Nicolai Wendt wurde im Jahr 2019 von Joachim Funke promoviert, nachdem er zuvor sein Student und wissenschaftlicher Mitarbeiter gewesen ist. Sie verbindet die Identifikation mit der theoretischen Psychologie, über die sie 2022 unter dem Titel *Wohin steuert die Psychologie?* gemeinsam ein programmatisches Buch aus philosophischer und psychologischer Sicht veröffentlicht haben. Als Denkpsychologe und als theoretischer Psychologe ist Alexander Wendt nicht nur in der Geisteshaltung der kritischen Wissenschaft, sondern auch persönlich von Joachim Funke geprägt worden.

Komplexität und Problemlösen sind Schlüsselwörter, die in der kognitions- und denkpsychologischen Forschung der letzten Jahrzehnte ihre Bedeutung entfaltet haben. Joachim Funkes psychologisches Werk ist mit beiden Begriffen eng verbunden. In seinem Denken über komplexes Problemlösen laufen verschiedene Forschungsrichtungen zusammen, etwa Handlungspsychologie, Systemtheorie und die computationale Theorie des Problemlösens. Der Band versammelt vielfältige Beiträge aus diesem Forschungszusammenhang, seien es biografische Kommentare, wissenschaftliche Essays oder Originalbeiträge.



**UNIVERSITÄT
HEIDELBERG**
ZUKUNFT
SEIT 1386

ISBN 978-3-96822-172-4



9 783968 221724