

# Die vier Wellen der psychometrischen Forschung zu komplexen Problemen

MATTHIAS STADLER

Ludwig-Maximilians-Universität München

**A**ls 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](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>