

# 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.<sup>1</sup> 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).

---

<sup>1</sup> 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  
 $\Sigma$ : Eingabealphabet  
 $\Omega$ : Ausgabealphabet  
 $\delta$ : Übergangsfunktion:  $\delta: Q \times \Sigma \rightarrow Q$   
 $\lambda$ : 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  $\Sigma$  besteht aus externen Ereignissen, das Ausgabealphabet  $\Omega$  aus Verhalten. Die Übergangsfunktion  $\delta$  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  $\Sigma$  (Cartesisches Produkt) eine Regel. Die Ausgabefunktion  $\lambda$  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  $\Sigma$ ) 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  $\delta$  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.

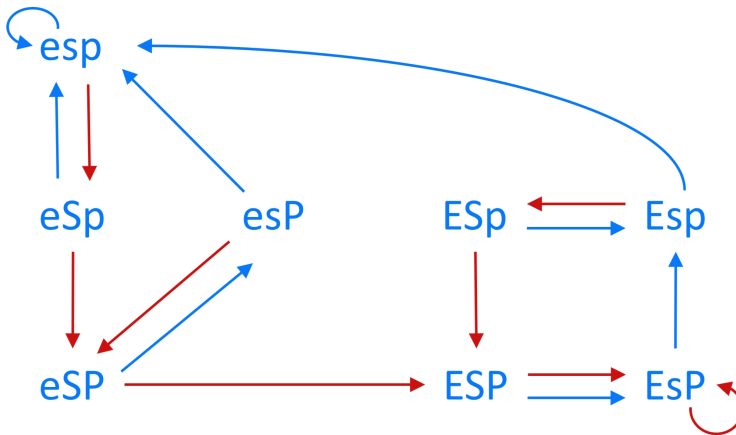
- Zu b)  $eSp$  ist der Zustand, in dem erfolgreich die geforderte SK angewendet wird; das geht eine Zeit lang gut, resultiert aber im Zustand  $eSP$ , der anzeigt, dass schon länger SK angewendet wurde.
- Zu c) Bei weiter geforderter SK kommt es zur Ego Depletion; ich nehme hier an, dass mit deren Auftreten 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.<sup>2</sup>

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  $\Delta$ , 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  $I$ , definiert nach den Konventionen des Dual-Processing-Ansatzes. Unterschieden wird zwischen Verarbeitung, die überwiegend Typ 1 zugeordnet ist, und Verarbeitung, die überwiegend Typ 2

---

<sup>2</sup> 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  $\Phi$  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.<sup>3</sup>

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

<sup>3</sup> 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  $\Delta$ . 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  $\Delta$ , Informationsverarbeitung  $I$  und Fertigkeiten  $\Phi$  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.<sup>4</sup>

## 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.

---

<sup>4</sup> 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](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](http://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>