

Komplexität und freier Wille

Anmerkungen zu Arbeiten von Stephen Wolfram und Daniel Dennett¹

Rainer Hammwöhner
Universität Regensburg
Informationswissenschaft

1. Einführung

Aus seinen Arbeiten über zelluläre Automaten leitet Stephen Wolfram Aussagen über Möglichkeit und Natur des freien Willens ab. In diesem Beitrag soll, soweit in diesem Kontext möglich, Wolframs Position mit ihren Voraussetzungen erläutert werden. Sie wird dann mit Daniel Dennetts Überlegungen zu diesem Thema in Beziehung gesetzt.

2001 erschien das lang erwartete, mehr als 1000 Seiten starke Großwerk von Stephen Wolfram, *A New Kind of Science (NKS)* (Wolfram, 2001). Die hohen, mit dem Werk verbundenen Erwartungen begründeten sich aus der Biographie des Verfassers. Stephen Wolfram ist der Entwickler von Mathematica². Schon im Alter von 15 Jahren publizierte er seinen ersten wissenschaftlichen Artikel, mit 20 promovierte er am *Californian Institute of Technology*. Der mit dem Vertrieb von Mathematica verbundene wirtschaftliche Erfolg erlaubte es Stephen Wolfram, über 20 Jahre völlig unabhängige Forschung auf dem Gebiet der zellulären Automaten und der Komplexitätstheorie zu betreiben, die dann durch die Publikation von *NKS* vorgestellt wurden. Das Buch wurde mehrfach besprochen³. Es zeigte sich jedoch unmittelbar, dass formale Fragen und deutliche Verstöße gegen wissenschaftliche Gepflogenheiten die Würdigung der wissenschaftlichen Ergebnisse erschwerten bzw. überschatteten. *NKS* wurde im Eigenverlag publiziert, wurde also nicht als Publikation eines Fachverlages, wie sonst üblich, zuvor begutachtet. Das Buch kreist ausschließlich um die Ergebnisse von Wolframs Forschung und vermeidet fast weitgehend, diese nachvollziehbar in ein wissenschaftliches Umfeld einzubetten. Ergänzt wird diese Haltung durch außerordentlich selbstbewusste Formulierungen zur Würdigung des eigenen Tuns. So fehlt in fast keiner Rezension der Hinweis, dass man dieses Buch nur bespreche, weil man einen Stephen Wolfram nicht einfach übergehen könne. Es gibt aber durchaus auch Rezensenten, die darauf hinweisen, dass hier in der Tat ein fundamental neuer Zugang zu vielen wissenschaftlichen Fragestellungen gefunden worden sei. Zu den Kernaussagen von *NKS* gehört, dass viele komplexe Erscheinungsformen in der Realität aus der rekursiven Anwendung einfacher Grundoperationen resultieren. Als eine der zentralen Aufgaben der Wissenschaft sieht es Wolfram an, zu studieren, empirisch festzustellen, welche Arten von Grundoperationen zu welcher Art von komplexen Phänomenen führen. Dass dies, so wie er es in den letzten 20 Jahren getan hat, durch Computersimulationen erforscht werden kann, sieht er darin begründet, dass alle in der Welt ablaufenden Prozesse im Prinzip als Rechenoperationen

¹Dieser Text ist erschienen in: Reinhard Rapp, Peter Sedlmeier, Gisela Zunker-Rapp (Hrsg.): Perspectives on Cognition. A Festschrift for Manfred Wetzler. Pabst, 2006, S. 459-470.



Dieser Text ist unter der folgenden Creative Commons Lizenz lizenziert: Attribution-NonCommercial-NoDerivs 2.0 Germany (<http://creativecommons.org/licenses/by-nc-nd/2.0/de/>).

² Mathematica ist eines der meistverkauften mathematischen Programmpakete mit Funktionen zu Algebra, Numerik, Grafvisualisierung (Mathematica 2005)

³ Eine Zusammenstellung von Rezensionen stellt Clark (2005) zur Verfügung.

aufgefasst werden können. Er nennt dies das Prinzip der *computational equivalence*⁴. Eine Anwendung seiner Überlegungen liegt für Wolfram darin, die Möglichkeit des freien Willens in einer deterministischen Welt zu erklären. Im folgenden werden die Kernaussagen zu diesem Thema, die über zahlreiche Fußnoten verstreut sind, extrahiert und zu anderen Modellen des freien Willens in Beziehung gesetzt.

2. Zelluläre Automaten

Zelluläre Automaten (Zellulärer Automat 2005) dienen der Modellierung diskreter dynamischer Systeme. Sie bestehen aus einer Menge von Zellen, auf der eine Nachbarschaftsbeziehung definiert ist. Eine Zustandsmenge definiert, welche Zustände die Zellen einnehmen können. Eine lokale Überföhrungsfunktion legt fest, wie sich der Folgezustand einer Zelle aus ihrem eigenen Zustand und aus den Zuständen der Zellen ihrer Nachbarschaft ergibt. Berechnungen zellulärer Automaten erfolgen, indem, ausgehend von einem initialen Zustand, in einer Folge von Schritten jeweils für alle Zellen neue Zustände berechnet werden.

Die Idee zellulärer Automaten wurde bereits in den 40er Jahren des 20. Jahrhunderts von Stanislaw Ulam und John von Neumann entwickelt. In den 70er Jahren erlangte Conway's *Game of Life* (Game of Life 2005) – ein auf zweidimensionalen zellulären Automaten beruhendes Simulationsmodell – Berühmtheit. Die Zellen sind hier in einer zweidimensionalen Matrix so angeordnet, dass jede Zelle über 8 unmittelbar benachbarte Knoten verfügt, deren Zustände in die Übergangsfunktion eingehen. Jede Zelle kann genau zwei Zustände einnehmen: an oder aus. Die Übergänge sind definiert wie folgt:

- Hat die Zelle genau 2 aktivierte Nachbarn, bleibt ihr Zustand unverändert.
- Hat sie genau 3 aktivierte Nachbarn, wird sie im nächsten Zustand aktiviert sein, unabhängig vom Vorgängerzustand.
- Hat sie weniger als 2 oder mehr als drei aktivierte Nachbarn, so ist sie im nächsten Zustand deaktiviert.

Aktiviert man die Simulation mit einem nicht-trivialen Anfangszustand, so entstehen sehr schnell komplexe Strukturen, unter denen statische, zyklisch bewegliche und wirklich dynamische zu unterscheiden sind. Unter letzteren sind bewegliche (Gleiter), strukturzeugende (Gleiterkanonen) und strukturvernichtende (Fresser) zu unterscheiden.

Wolfram arbeitet in seinen Simulationen primär mit eindimensionalen zellulären Automaten⁵ – diese repräsentieren sozusagen eine zweidimensionale Raum-Zeit. Er untersucht dabei, wie unterschiedliche Übergangsfunktionen das Verhalten der Automaten über eine Vielzahl von Generationen hinweg beeinflussen, wenn von einem Zufallsmuster ausgegangen wird. Grundlage sind dabei die zweidimensionalen Raum-Zeit-Muster, die Abfolge der 1eindimensionalen Zustandsräume. Es ergibt sich das Resultat, dass sich vier Klassen von Automaten unterscheiden lassen (Wolfram, 2001: 231 ff.), die hier näher beschrieben werden (s. Abb. 1):

1. Die Automaten der Klasse eins zeigen ein einfaches Verhalten, fast alle Ausgangszustände föhren zu einem uniformen Endzustand.

⁴ Er folgt darin einer Idee, die schon Konrad Zuse ausgearbeitet hat (1969).

⁵ Er geht davon aus, dass höherdimensionale Automaten keinen höheren Grad an Komplexität aufweisen (Wolfram, 2001: 169 ff.).

2. Bei Exemplaren der Klasse zwei gibt es eine Vielzahl von möglichen Endzuständen, die aber aus simplen unbegrenzten oder repetitiven Strukturen bestehen.
3. Automaten der dritten Klasse zeigen ein komplexeres Verhalten. Es entstehen Muster, die den Anschein des Zufalls erwecken, dennoch aber einfache Teilmuster, wie Dreiecke oder andere kleinräumige geometrische Muster enthalten.
4. In der vierten Klasse zeigen die Automaten im Verhalten eine Kombination von Zufälligkeit und Ordnung. Lokale einfache Strukturen entstehen, die sich im Raum bewegen und auf komplexe Art und Weise miteinander interagieren.

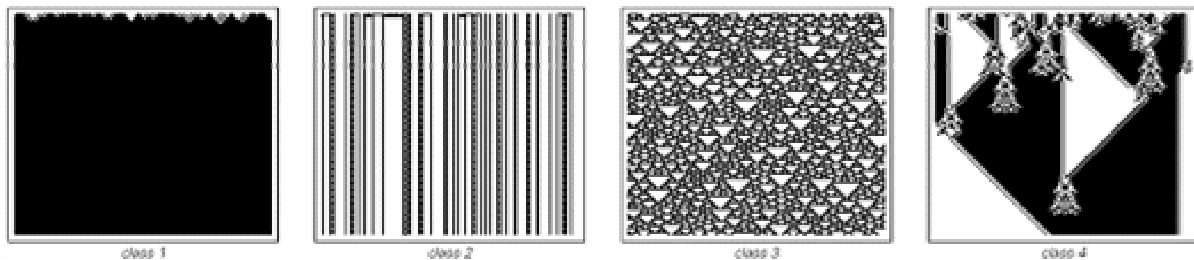


Abbildung 1 – Beispiele für die 4 Klassen zellulärer Automaten⁶ (Wolfram, 2001: 231)

Es ist zu betonen, dass die Klassifikation der Automaten rein empirisch-deskriptiv ist. Für eine Klassifikation der Automaten anhand ihrer Übergangsfunktion existiert keine Theorie. Dieser empirische Ansatz in Bezug auf Fragen der Komplexität und Berechenbarkeit zählt zu den Besonderheiten von *NKS*. Wir wollen im folgenden einige der Kernthesen von *NKS* herausgreifen, die für unsere weitere Argumentation von Bedeutung sind.

2.1 Universelle zelluläre Automaten

Für einen Automatentyp – denjenigen mit der Übergangsfunktion (Regel) 110⁷ (s. Abb. 2) – sind nun detailliertere Untersuchungen angestellt worden. Es ließ sich zeigen, dass dieser Automat in der Lage ist, universelle Berechnungen auszuführen. Über den Umweg über zyklische Tag-Systeme konnte bewiesen werden, dass Regel 110 alle Operationen einer universellen Turing-Maschine ausführen (Wolfram, 2001: 675 ff.), und somit – nach der *Church-Turing-These* – jede berechenbare Funktion berechnen kann. Ähnliche Ergebnisse wurden schon für andere zelluläre Automaten aufgezeigt. Die Besonderheit in diesem Fall ist die Einfachheit von Regel 110. Regel 110 ermittelt den Nachfolgezustand einer Zelle allein aufgrund ihres Ausgangszustands und des Zustands der unmittelbar benachbarten Zellen. Es sind also 2^3 mögliche Übergänge zu definieren. Wolfram sieht in der Einfachheit dieses universellen Automaten einen Hinweis auf sein Prinzip der *computational equivalence*. Weiterhin vermutet er, dass alle Automaten der Klasse vier diese Eigenschaft aufweisen. Beweise stehen hier aber noch aus.

⁶ Die Abbildungen aus diesem Abschnitt entstammen *NKS* (Wolfram, 2001). Sie wurden von Wolfram Media unter www.wolframscience.com (zitiert am 19.5.2005) online zur Verfügung gestellt.

⁷ Wolfram hat mit einer Vielzahl von Übergangsfunktionen experimentiert, die durchnummeriert wurden.



Abbildung 2 – Regel 110⁸



Die ersten fünf Schritte von Regel 110

2.2 Irreduzible Berechnungen

Aus dem beobachteten Verhalten zahlreicher komplexer zellulärer Automaten, hat Wolfram die Hypothese abgeleitet, dass es kein Verfahren gibt, die zukünftigen Zustände eines solchen Automaten mit einem geringeren Aufwand zu prognostizieren, als diesen Automaten seine Berechnungen selbst ausführen zu lassen. Diese Hypothese ist von Bedeutung, nicht nur für die Theorie zellulärer Automaten, sondern – akzeptiert man das Prinzip der *computational equivalence* (s.u.) – auch für die Naturwissenschaften. Sind nämlich in der Natur äquivalente, irreduzible Prozesse zu verzeichnen, so sind damit weitgehende Beschränkungen für experimentelle Vorhersagen gegeben. In einer solchen Welt könnte nicht einmal ein *Laplacescher Dämon* bei vollständiger Kenntnis des Ausgangszustands und der Naturgesetze – bei Laplace als Bewegungsgesetze zu denken – eine Prognose über zukünftige Zustände des Systems mit Sicherheit leisten⁹.

2.3 Entscheidbarkeit

Die Frage nach der Entscheidbarkeit eines Problems beinhaltet die Frage nach der Existenz eines Algorithmus, der in einer endlichen Zahl von Schritten eine Antwort auf dieses Problem ermittelt. Im Zusammenhang mit eindimensionalen zellulären Automaten sind alle Fragen nach dem Grenzverhalten (für eine unbegrenzte Anzahl von Schritten) unentscheidbar. Das heißt, ein solcher Algorithmus existiert nicht (Wolfram, 2001:1138). Fragen hinsichtlich des Verhaltens über eine endliche Zahl von Schritten sind zumeist entscheidbar, im Gegensatz zu zweidimensionalen Automaten. Die durch die Anwendung der einfachen Grundregeln des *Game of Life* entstehenden Strukturen und Strukturwandlungen sind so geartet, dass für zwei beliebige vorgegebene Strukturen nicht entscheidbar ist, ob sie durch wiederholte Anwendung dieser Regeln ineinander überführbar sind. Dies ist ein Analogon zum Halteproblem (Sutner, 2004)¹⁰. Dies ergibt sich aus der nachgewiesenen Universalität des *Game of Life* und anderer unendlicher zellulärer Automaten und gilt demnach auch für Regel 110.

Für finite zelluläre Automaten sind diese Fragen entscheidbar. Sie fallen für 2-dimensionale zelluläre Automaten in die Klasse der NP-vollständigen Probleme.

2.4 Prinzip der computational equivalence

Die von Wolfram formulierte These (Wolfram 2001: 715 ff.) kann in zwei Teilthesen zerlegt werden, die einzeln begründbar sind.

⁸ Die Abbildung zeigt die möglichen Zustandsübergänge einer Automatenzelle in Abhängigkeit von den Zuständen ihrer beiden Nachbarzellen.

⁹ Wenn man die Möglichkeit ausschließt, dass der Dämon ein vollständiges Abbild des Universums „schneller ausführen“ könne. Da dieses offensichtlich nicht innerhalb dieses Universums stattfinden könnte und da man dem Abbilduniversum aufgrund der Vollständigkeit der Abbildung keinen reduzierten Realitätsgehalt zuweisen könnte, wäre in diesem Fall auch weniger von einer Prognose als von einer Beobachtung zu sprechen. Auch Dennett (2003:28) bezieht sich in seinen Ausführungen auf diese Interpretation des Laplaceschen Dämons.

¹⁰ Auch Dennett erwähnt die Universalität des *Game of Life* (Dennett, 2003: 46), weist aber auf mögliche formale Konsequenzen nur sehr summarisch hin (Dennett, 2003: 51).

- a. Alle in der Welt auftretenden Prozesse, seien sie natürlich oder künstlich hervorgebracht, sind hinsichtlich ihrer Komplexität äquivalent. Selbst scheinbar einfachen Phänomenen liegt die höchstmögliche Berechnungskomplexität zu Grunde.
- b. Alle oben genannten Prozesse entsprechen Berechnungsprozessen, die auch auf universellen Turingmaschinen ausführbar sind.

Zunächst bleibt festzustellen, dass beide Thesen empirischer Natur sind. Sie sind falsifizierbar. Nur die erstere wird von Wolfram in seinem Text explizit behauptet, die zweite ergibt sich hingegen implizit. Aus dieser These leitet Wolfram weitgehende Konsequenzen für Mathematik und Naturwissenschaften ab. Für uns ist relevant, dass die These über irreduzible Berechnungen so gestützt wird. Falls die Mehrfachanwendung einfacher Regeln schon die höchste Komplexitätsklasse erreicht, so ist nicht ersichtlich, wie eine prognostische Berechnung möglich sein sollte.

These *b* nimmt eine zentrale Rolle bei Plausibilitätsbetrachtungen zu These *a* ein. Wolfram zeigt auf, dass universelle Berechnungen schon mit sehr einfachen Strukturen (z.B. Regel 110) möglich sind. Dies fasst er als Hinweis auf, dass Universalität in der Natur verbreitet ist. Als weiteres Indiz führt er die erfolgreiche Simulation biologischer Prozesse mit Hilfe zellulärer Automaten an. Die Äquivalenzbehauptung ist aber nur dann auf diese Weise zu stützen, wenn man annimmt, dass in der Natur keine Berechnungsvorgänge vorkommen, die von Turingmaschinen nicht ausgeführt oder simuliert werden können. Die Frage, wie sich Turing-Berechenbarkeit zu physikalischen Prozessen verhält, wurde schon von Chaitin (82) aufgeworfen. Copeland (04) warnt vor voreiligen Schlüssen – etwa durch unzulässige Verallgemeinerungen der Church-Turing-These –, und sieht bisher keine zwingenden Argumente für diese Analogie. Penrose geht sogar von der Existenz determinierter, aber nicht berechenbarer Prozesse in der Natur aus – im Zusammenhang mit der postulierten Quantengravitation (91:421). Alle Argumente, die in Richtung auf die Existenz nicht durch Turing-Maschinen berechenbarer Funktionen in der Natur abzielen, würden allerdings die These der Berechnungsirreduzibilität für die involvierten Prozesse eher stärken als schwächen.

3. Freier Wille

Im folgenden Kapitel werden die aus *NKS* ableitbaren Implikationen für das Konzept des freien Willens zusammengestellt und mit anderen Auffassungen zu dieser Problematik in Beziehung gesetzt.

3.1 Freier Wille in *NKS*

Obwohl im einleitenden ersten Kapitel ausdrücklich auf die philosophischen Implikationen der *NKS* hingewiesen wird und unter diesen das Problem des freien Willens explizit erwähnt ist (Wolfram, 2001: 10), nimmt die Diskussion dieser Fragestellung einen vergleichsweise geringen Raum ein.

Eine zentrale Rolle in Wolframs Argumentation nimmt die Hypothese der Berechnungsirreduzibilität ein (Wolfram, 2001: 750 ff.). Ein System, dessen Verhalten keinem einfachen Algorithmus folgt, das also nicht „auszurechnen“ ist, kann als freier angesehen werden, als ein System, das diese Eigenschaft nicht aufweist. Wolfram sieht Freiheit hier also als ein graduelles Phänomen an, für das aus der Berechnungskomplexität hinreichende Bedingungen ableitbar sind. Den Konflikt von Vorwissen und freiem Willen

sieht Wolfram schon bei Augustinus angedeutet¹¹ (Wolfram, 2001:1136). Aber auch Wittgenstein sieht einen Widerspruch zwischen Vorhersagbarkeit und freiem Willen (Dilman 99, S. 248 f.). Nicht-Vorhersehbarkeit wird von vorgenannten Autoren aber als notwendige Bedingung für freien Willen angesehen.

Wir beobachten in der Natur Verhaltensweisen – sowohl von Tieren als auch von Menschen –, die eine komplexe Kombination von wiederkehrenden Mustern und Unvorhersehbarem enthalten. Diese Verhaltensstruktur ist konsistent mit dem von Wolfram formulierten Modell (Wolfram, 2001: 1135). Schon einfache zelluläre Automaten zeigen, dass simple Regeln auf der Ebene der Zellen ein nicht kalkulierbares Verhalten auf der Ebene der zeitlichen Evolution des Gesamtautomaten nicht ausschließen. Es ist also nicht möglich, den auf der Mikroebene zu konstatierenden Determinismus unreflektiert auf die makroskopische Ebene zu übertragen. Aus dem Determinismus der Mikroebene folgt keine Vorhersehbarkeit auf der Makroebene. Vielmehr ist das Verhalten des Automaten nicht auf isolierbare Einzelfaktoren – Ausgangsbedingungen, Übergangsregeln – zurückzuführen. Das Verhalten ergibt sich erst aus der Gesamtheit der internen Zustände im Kontext der jeweiligen Umwelt. Somit kann jede Handlung nur einem System als Ganzem – auch im Sinne einer Verantwortung – zugeschrieben werden (Wolfram, 2001: 1136).

3.2 Philosophische Grundpositionen zum Thema

Die Frage nach der Willensfreiheit ist die Frage danach, ob der Determinismus der naturwissenschaftlichen Erklärungsmodelle und die den Entscheidungen des Menschen unterstellte Freiheit in einem unauflösbaren Widerspruch stehen und wie diese gegensätzlichen Phänomene gegebenenfalls nebeneinander bestehen können. Determinismus bedeutet in diesem Kontext, dass alle mentalen Zustände und Handlungen durch vorgängige Ursachen notwendig hervorgerufen werden (Weatherford, 2005). Ausgangspunkt und Maßstab der Überlegungen zum Begriff der Freiheit ist die Alltagssprache, die sich auf die kleinen und großen Lebensentscheidungen des Menschen bezieht. Die philosophischen Freiheitsmodelle implizieren jeweils mehr oder weniger große Bedeutungsverschiebungen.

In dieser Frage sind grundsätzlich zwei Positionen möglich. Entweder sind freier Wille und Determinismus miteinander vereinbar – dies nennt man die kompatibilistische Auffassung – oder nicht – dies ist der Inkompatibilismus, mit dem wir uns zunächst befassen wollen. Als Antwort auf den Widerspruch kann entweder der freie Wille verworfen werden – diese Position wird als harter Determinismus bezeichnet – oder der Determinismus – dies ist die libertarische Position.

Roderick Chisholm (1964) als immer noch sehr prominenter Vertreter dieser Teilströmung des Inkompatibilismus fordert, dass freie Entscheidungen nur dann gegeben seien, wenn der Handelnde unverursacht, als „unbewegter Beweger“ eine völlig neue Kausalkette initiiere. Die Kritik an dieser Agentenkausalität richtet sich primär dagegen, dass Handlungen so völlig von den Motiven und Präferenzen einer Person abgelöst würden. Würde Motiven aber ein Einfluss zugestanden, so könnte nicht vom Beginn einer neuen Kausalkette die Rede sein.

Peter van Inwagen (1983) hingegen betont die Existenz geschlossener Kausalketten von Beginn des Universums an. Da wir aber weder Einfluss auf die Vergangenheit haben, die Ausgangsbedingungen für jede Handlung festlegt, noch auf die Naturgesetze, welche die Art

¹¹ Er versäumt aber, darauf hinzuweisen, dass Augustinus den Konflikt als aufgelöst ansieht. „Weil Gott auch unseren zukünftigen Willen vorherweiss, folgt daraus trotzdem keineswegs, dass wir nicht mit unserem Willen etwas wollen“ (Augustinus, 1961: 123). Vielmehr sieht Augustinus die göttliche Vorsehung sogar als eine Voraussetzung unseres freien Willens an (Augustinus, 1961: 124).

der Kausalbeziehungen festlegen, können wir den Ablauf der Dinge nicht beeinflussen. Jeglicher Eindruck von Freiheit in unseren Handlungen muss also Illusion sein.

Ein Kompatibilist, der es für möglich hält, dass Menschen trotz eines grundsätzlich gegebenen Determinismus zumindest manchmal oder unter bestimmten Bedingungen frei in ihren Entscheidungen und Handlungen sein können, wird zunächst die Prämissen des harten Deterministen angreifen. Eine Handlung ist – so die kompatibilistische Diktion – dann frei, wenn der Akteur anders hätte handeln können, wenn er sich anders entschieden hätte. Die Analyse des hier verwendeten Möglichkeitsbegriffs und der Handlungsalternativen vor einem deterministischen Hintergrund nimmt in kompatibilistischen Ausführungen (z.B. Bok, 1998) zumeist einigen Raum ein.

Eine andere, derzeit prominente Argumentationslinie verknüpft Willensfreiheit mit der Reflexionsfähigkeit des handelnden Subjekts (Frankfurt, 1971). Während schon Tiere Wünsche, Bestrebungen und Neigungen haben – etwa zu fressen oder zu spielen – kann nur der Mensch eine reflektierende Position zu seinem Willen einnehmen. Er kann seinen Willen annehmen oder ablehnen, er kann also Willensakte höherer Ordnung vollbringen. Ein Altruist kann seinen Willen zur guten Tat selbst wollen, während ein Trinker seinen eigenen Willen zum Alkoholkonsum auf einer höheren Ebene ablehnen kann. Willensfreiheit ist dann gegeben, wenn zwischen den Willensakten unterschiedlicher Ebenen – wie bei dem eben erwähnten Altruisten – kein Widerspruch besteht.

Pauen wiederum stellt die personale Integrität und Autonomie in den Vordergrund (2004). Seine Minimalanforderungen an personale Freiheit besagen, dass Handlungen nicht allein auf externe Ursachen zurückzuführen sein dürfen (Autonomieprinzip), dass sie aber auch nicht zufällig sein dürfen (Urheberprinzip). Letzteres wird dann als gewährleistet angesehen, wenn sie mit persönlichen Präferenzen, das sind Grundhaltungen, die im Verlauf des Lebens erworben wurden, übereinstimmen.

Wir haben nunmehr einige Auffassungen zur Natur des freien Willens kurz vorgestellt. Es scheint zwischen diesen Überlegungen und der von Wolfram vertretenen These keine unmittelbare Verbindung gegeben. Die Frage der Vorhersehbarkeit, die bei Wolfram eine so zentrale Stellung einnimmt, spielt bei den anderen Autoren keine Rolle. Bei Wolfram wiederum sind keine Hinweise auf eine Abgrenzung zwischen handelndem Agenten und seiner Umwelt gegeben, intentionale Zustände sind nicht angesprochen. Es ist also zunächst eine Brücke zwischen diesen prima facie inkompatiblen Betrachtungsebenen zu konstruieren. Der größere Teil der erforderlichen Überlegungen wurde schon von Daniel Dennett geleistet, dessen Ausführungen wir im folgenden kurz rekapitulieren wollen.

3.3 Dennett über Freiheit, alternative Möglichkeiten und Kausalität

Dennetts Ziel ist es, zu zeigen, dass die Freiheit von Lebewesen ein Resultat der Evolution ist (2003). Zunehmende Freiheit zeigt sich in einem wachsenden Repertoire an Verhaltensformen, mit denen ein Organismus auf komplexe Situationen reagieren kann, die sogar in Antizipation zukünftiger Situationen eingesetzt werden können. Diese Freiheit wird ermöglicht durch die immer komplexer werdende Struktur der Organismen, so dass ein Überwiegen interner Prozesse über externe Ursachen möglich wird (vgl. auch die von Pauen definierten Kriterien).

Auch Dennett sieht sich vor der Herausforderung, zu erläutern, wie sein Konzept von Freiheit mit dem Determinismus der physischen Welt koexistieren kann. Er holt hier zu einer umfangreichen Begründung aus, die eine Analyse der Vorstellungen von alternativen Möglichkeiten und von Kausalität einschließt. Interessant wird seine Argumentation vor allem durch die Struktur des Modells, das er ihr zu Grunde legt.

Zunächst legt Dennett eine Grundlage für weitere Gedankenexperimente, indem er sich ein einfaches Weltmodell schafft (2003: 25-63). In einem ersten Schritt geht er von einem dreidimensional euklidischen Demokrit-Universum aus, dessen Atome unteilbar sind und Raumzellen gleichen Volumens – Voxels – einnehmen¹². Dieses Modell benutzt Dennett zunächst, um den Zustandsraum einer derartigen Struktur zu verdeutlichen, die beliebig mit Atomen belegt sein kann. Eine Raum-Zeit kann dann als Abfolge solcher Zustände gedacht werden. Diese Zustandsfolgen, charakterisieren dann deterministische (wenn Regularitäten auftreten), indeterministische (wenn probabilistische Zusammenhänge bestehen) und nihilistische (wenn keinerlei Regularitäten bestehen) Universen. Dennett verfolgt diese Idee aber nicht weiter. Vielmehr nimmt er, bevor er zu den eigentlichen Überlegungen voranschreitet, eine weitere Vereinfachung vor. Als Modell wählt er im folgenden das *Game of Life*. Durch die Wahl eines zellulären Automaten als Modell ist nunmehr eine Anschlussmöglichkeit an die oben erläuterten Forschungsergebnisse gegeben.

In seiner Untersuchung des *Game of Life*-Universums unterscheidet Dennett zunächst zwei Ebenen:

- Auf der untersten Ebene – Dennett nennt sie physikalisch – ist nur eine Verteilung von ein- oder ausgeschalteten Raumpunkten zu beobachten. Als Ereignisse sind der Zustandswechsel von Zellen zu verzeichnen.
- Auf der Designebene treten Objekte (Glider, Eater) auf, komplexe Muster, die über die Zeit eine gewisse Persistenz aufweisen. Sie können sich bewegen und auch ihre Gestalt verändern.

Jede dieser Ebenen erfordert zur Beschreibung ihre eigene Ontologie. Dennett setzt sein Gedankenspiel fort und erwägt die Möglichkeit, dass in einem riesigen *Game of Life* Objekte entstehen könnten, die so evolvieren, dass sie ein auf ihre „Umwelt“ gerichtetes Verhalten an den Tag legen können¹³. Damit wäre dann eine dritte, intentionale Betrachtungsebene gegeben. Man könnte den Objekten Wahrnehmungen, Pläne, Ziele o.Ä. zuschreiben. Dennett ordnet nun jede Redeweise über alternative Möglichkeiten, die Voraussetzung einer Debatte über freien Willen sind, zumindest der Design-Ebene zu, während die deterministische Physik der untersten Ebene verhaftet bleibt. Insofern bestehe kein Widerspruch zwischen Freiheit und Determinismus. Dieses Gedankenexperiment ist weitgehend konform zu den Resultaten von Wolframs Experimenten. Zwar vermeidet letzterer eine Vergegenständlichung der beobachteten Muster. Er betont aber auch, dass Zellebene und übergeordnete Musterebene letztlich inkommensurabel sind, dass aus den Eingangszuständen und den Übergangsregeln kein Schluss auf die entstehenden Muster möglich ist.

Nachdem eine Eingangsplausibilität hergestellt ist, präzisiert Dennett den Begriff der Kausalität bzw. der Verursachung (2003: 70 ff.)¹⁴. Zunächst stellt er eine Liste von Eigenschaften zusammen, die von Verursachungsbeziehungen geteilt werden. Dies sind keine harten Kriterien¹⁵, die gegeben sein müssen, will man von Verursachung sprechen. Vielmehr tragen sie alle zur Evidenz für das Vorhandensein einer Verursachung bei. Ihr Beitrag ist dabei unterschiedlich gewichtig. Sie sind hier nach absteigender Bedeutung geordnet. In den

¹² Er führt hier etwas ausführlicher eine Idee von W.v.O. Quine aus (1969: 147).

¹³ Hinter diesem Gedankenspiel steht – bei erweiterter Methodik – die Forschungsrichtung des Artificial Life (Bedau 2004).

¹⁴ Eine formalere Fassung dieser Argumentation, die modallogisch fundiert ist, findet sich in (Taylor / Dennett, 2002). Dort fehlt allerdings der Bezug zu zellulären Automaten.

¹⁵ Diese Kriterien entsprechen einem weitgehendem Konsens in der Forschung zur Kausalität, vgl. z.B. (Mackie 2004).

Kriterien werden folgende Kürzel verwendet: Ereignis B wird von A in einer Welt x verursacht. Dabei ist X eine Menge von Welten, die ähnlich zu x sind, wobei die Ähnlichkeitsrelation nicht näher spezifiziert ist. Eingeschlossen sollen jedenfalls jeweils einige Welten sein, in denen A und B eintreten, nur eines von beiden oder keines der Ereignisse zu beobachten sind.

Kausale Notwendigkeit ist gegeben, wenn in allen Welten in X , in denen B eintritt auch A eingetreten ist.

Kausal hinreichend ist A , wenn jedes Mal in X , wenn A eintritt auch B eintritt.

Unabhängigkeit: A und B sollen in keinem logischen Zusammenhang stehen.

Zeitliche Reihenfolge: A sollte sich vor B ereignen.

Dennett verwendet zur Veranschaulichung dieser Kriterien Beispiele aus der Alltagswelt. Leider greift er nicht wieder auf das *Game of Life* zurück. Hätte er dies getan, wären manche Probleme, die den obigen Kriterien implizit sind, offensichtlich geworden.

Zunächst hätte sich die Frage nach einer Präzisierung der Ähnlichkeitsrelation dringlicher gestellt. Wann sind zelluläre Automaten hinlänglich ähnlich? Vermutlich sollte die gleiche Übergangsfunktion gelten. Muss die Ausdehnung des Zellraums gleich groß sein? Um wie viele und welche Zellbelegungen dürfen sich die beiden Automaten zu welchen Zeiten unterscheiden? Darf das Verhalten des Automaten eine Rolle spielen? Macht man sich dann nicht eines Zirkelschlusses schuldig?

Die obige Definition der Kausalität geht, im Lichte von Wolframs Ausführungen, von problematischen Prämissen aus. Können wir überhaupt sinnvolle Annahmen über Ereignisfolgen in kontrafaktischen Welten treffen? Wie wir bereits festgestellt haben, ist die Frage nach der Überführbarkeit von einem Automatenzustand A in einen gegebenen anderen Zustand B für universelle zelluläre Automaten – wie das *Game of Life* – nicht entscheidbar. Dies trifft zwar nur für infinite Automaten¹⁶ zu. Das Prinzip der Berechnungsirreduzibilität verhindert dennoch, Kalkulationen über alternative Möglichkeiten mit der Sicherheit vorzunehmen, wie sie für eine Entscheidung über die obigen Kriterien erforderlich wäre. Somit scheint die Frage nach der Willensfreiheit in dem zu ihrer Klärung von Dennett vorgeschlagenen Modell nicht entscheidbar zu sein.

Die Forderung nach logischer Unabhängigkeit ist zunächst sehr plausibel. Als Beispiel führt Dennett an „Mary’s singing and dancing caused her to sing and to dance“. Hier wird in der Tat kaum jemand eine Kausalbeziehung erkennen. Dennett verallgemeinert so, dass zumindest eine Welt existieren müsse, in der nur A oder nur B gelte. Auf diese Weise könnten auch Sätze ausgeschlossen werden wie „ $1+1=2$ verursacht $2+2=4$ “. Letzteres wäre eigentlich leichter zu haben. Weder „ $1+1=2$ “ noch „ $2+2=4$ “ sind Ereignisse. Also können keine Verursachungsbeziehungen zwischen ihnen bestehen. Auf den zweiten Blick ist das Urteil vielleicht nicht mehr so sicher. Die Operationen zellulärer Automaten sind ebenso rein formal. Im einfachsten Fall würden sich zelluläre Automaten hiermit als ungeeignet für Gedankenexperimente über Kausalität und freien Willen herausstellen. Gilt Wolframs These der „*computational equivalence*“, finden wir in der Realität auch nichts anderes vor als das Resultat von Berechnungen, wengleich auch Wolfram nie behauptet hat, die Welt *sei* ein zellulärer Automat.

¹⁶ Zur Komplexität zellulärer Automaten vgl. auch (Culik et al. 90).

4. Schlussfolgerungen

Wie Dennett in seinen Ausführungen durchaus plausibel machen kann, gibt es gute Gründe, von einer Vereinbarkeit von Determinismus und freiem Willen auszugehen. Gleichwohl wird deutlich, dass nicht nur praktische, sondern formale Gründe sich einem sicheren Wissen in dieser Frage frühzeitig entgegenstellen. Soviel haben wir aus den Arbeiten zu zellulären Automaten entnehmen können. Wolframs eigener Vorschlag, freien Willen mit algorithmischer Irreduzibilität gleich zu setzen, entbehrt aber leider der gleichen Überzeugungskraft. Worauf schon Ray Kurzweil (2001) hinwies: in Wolframs Modell müsste man auch einer Staubwolke freien Willen zusprechen. Hier zieht Wolfram Schlussfolgerungen, die durch die doch vergleichsweise einfachen Muster seiner Automaten der Klasse 4 (noch) nicht gedeckt werden. Sie mögen zwar berechnungsuniversal sein, aber auch auf Turingmaschinen können einfache und komplexe Programme definiert sein. Die Leistung liegt dann in der Komplexität des Programms und nicht in derjenigen der Turingmaschine an sich. Die Annahme von nicht turing-berechenbaren „Hypercomputations“ im Zusammenhang mit der Quantengravitation, von denen Penrose (1991, S. 421) ausgeht, zielen in die gleiche Richtung – der Determinismus soll von der Vorhersehbarkeit abgekoppelt werden, um freien Willen vorstellbar zu machen. Hier ist – im Gegensatz zu Wolframs Arbeiten – der ganze theoretische Rahmen noch nicht ausgeführt.

Die Frage nach Existenz und Natur des freien Willens könnte jedoch zu jenen gehören, die prinzipiell unser Erkenntnisvermögen übersteigen, so McGinn aus der Perspektive des transzendentalen Naturalismus (1996: 130-150): „Wir begreifen einfach nicht, in welcher Beziehung die Notwendigkeit der Materie zu den Möglichkeiten des Willens stehen, ...“ (1996: 147). Es ist aber noch nicht ausgemacht, ob diese Fragestellungen prinzipiell außerhalb unseres Erkenntnisvermögens liegen oder ob bisher nur die angemessenen Methoden und Begrifflichkeiten fehlen. Die Frage nach dem Verhältnis von Automatenzustand und Übergangsfunktion einerseits und Designebene bei Automaten andererseits, hat eine ähnliche Anmutung. Wolframs Arbeiten suggerieren hier eine dritte Antwortmöglichkeit. Aufgrund der algorithmischen Irreduzibilität könnte sich die Frage nach der Existenz der Willensfreiheit – zumindest in der von Dennett vorgeschlagenen Interpretation – als grundsätzlich und unabhängig von unserem Erkenntnisvermögen unbeantwortbar erweisen. Hier befinden wir uns aber bislang auf der Ebene von Hypothesen. Wolfram hat hier neue Perspektiven auf alte Fragen eröffnet. Darin dürfte sein Verdienst liegen. Ob damit die Qualifikation als „A New Kind of Science“ gerechtfertigt erscheint, muss hier nicht entschieden werden.

Die Ergebnisse, die wir hier in Bezug auf zelluläre Automaten erzielt haben, sollten unschwer auf neuronale Netze übertragbar sein. Auch hier sind Annahmen hinsichtlich der Universalität und Irreduzibilität sinnvoll. Wolfram vertritt in NKS jedenfalls diese Meinung (2001:1136). Von diesem Standpunkt aus gesehen erscheinen alarmistische Meldungen, wie sie etwa von Singer (2004) vertreten werden, dass neurowissenschaftliche Forschung die Existenz freien Willens widerlege, zumindest als verfrüht. Noch besteht kein Anlass, unsere alltägliche Redeweise von der Freiheit unserer Entscheidungen fallen zu lassen.

Danksagung

Meinen Kollegen Hans Rott und Christian Wolff sowie meiner Frau Christiane danke ich für ihre wertvollen Hinweise zu früheren Fassungen dieses Beitrags.

Besonderen Danke schulde ich Manfred Wetzler für die Kooperation während der gemeinsamen Zeit in Konstanz. Immerhin hat er durch die Zweitbetreuung meiner Dissertation eine der Voraussetzungen dafür geschaffen, dass ich heute an Veröffentlichungsprojekten wie diesem teilnehmen kann.

Lieber Manfred, ich wünsche Dir für die Zukunft einen Zugewinn an persönlicher Freiheit und die Möglichkeit damit zu tun, was Du willst.

Literaturangaben

- Augustinus (1961) *Der freie Wille*. Paderborn: Ferdinand Schöningh.
- Bedau, M.A. (2004) Artificial Life. In Floridi, L. (ed.), *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford: Blackwell Publishing, 197-211.
- Bok, H. (1998) Freedom and Practical Reason. Zitiert nach Watson, G. (2003, ed.), *Free Will*. Oxford, New York: Oxford University Press, 130-166.
- Chaitin, G.J. (1982) Gödel's Theorem and Information. In *International Journal of Theoretical Physics*, 22, 941-954.
- Chisholm, R. M. (1964) Human Freedom and the Self. Zitiert nach Watson, G. (2003, ed.), *Free Will*. Oxford, New York: Oxford University Press, 26-37.
- Clark, W. E. (2005) A Collection of Reviews of ANKOS and Links to Related Work, http://www.math.usf.edu/~eclark/ANKOS_reviews.html, zitiert am 19.5. 2005
- Copeland, B.J. (2004) Computation. In *The Blackwell Guide to the Philosophy of Computing and Information*, Oxford: Blackwell Publishing, 3-17.
- Culik II, K.; Hurd, L.P.; Yu, S. (1990) Computation Theoretic Aspects of Cellular Automata. In *Physica*, D45, 357-376.
- Dennett, D.C. (2003) *Freedom Evolves*. London: Penguin Books.
- Dilman, Ilman (1999) Free Will. An Historical and Philosophical Introduction. London, New York: Routledge.
- Frankfurt, H.G. (1971) Freedom of the Will and the Concept of a Person. Zitiert nach Watson, G. (2003, ed.), *Free Will*. Oxford, New York: Oxford University Press., 322- 336
- Game of Life (2005, 5.4.) *Wikipedia: The Free Encyclopedia*, http://de.wikipedia.org/wiki/Game_of_Life.
- Kurzweil, R. (2001) *Reflections on Stephen Wolfram's 'A New Kind of Science'*. <http://www.kurzweilai.net/articles/art0464.html?printable=1> (zitiert am 22.5.2005)
- Mackie, P. (2004) Causality. In Honderich, T. (ed.), *The Oxford Companion to Philosophy*. Oxford, New York: Oxford University Press, 131-133.
- Mathematica (2005, 13.6.) *Wikipedia: The Free Encyclopedia*, <http://de.wikipedia.org/wiki/Mathematica>.
- McGinn, C. (1996) *Die Grenzen vernünftigen Fragens*. Grundprobleme der Philosophie. Klett-Cotta.
- Pauen, M. (2004) *Illusion Freiheit. Mögliche und unmögliche Konsequenzen der Hirnforschung*. S. Fischer.
- Penrose, R. (1991) *Computerdenken*. Heidelberg: Spektrum.
- Quine, W.v.O. (1969) Propositional Objects. In Quine, W.v.O.: *Ontological Relativity and Other Essays*. Boston: Columbia University Press, 139-160.
- Singer, W. (2004) Verschaltungen legen uns fest: Wir sollten aufhören von Freiheit zu sprechen. In Geyer, C. (ed.), *Hirnforschung und Willensfreiheit. Zur Deutung der neuesten Experimente*. Frankfurt a.M.: Suhrkamp. 30-65.

- Sutner, K. (1995): On the Computational Complexity of Finite Cellular Automata. In *Journal of Computer and System Sciences*, 20, 87-97.
- Sutner, K. (2004) Computing in Cellular Automata. Report Carnegie Mellon University. <http://www-2.cs.cmu.edu/afs/andrew.cmu.edu/course/15/354/www/Lectures/ca-computation.pdf> (zitiert am 22.5. 2005).
- Taylor, C. und Dennett, D.C. (2001) Who's Afraid of Determinism? Rethinking Causes and Possibilities. In Kane, R.(ed.), *The Oxford Handbook of Free Will*. Oxford, New York: Oxford University Press, S. 257-277.
- Van Inwagen, P. (1983) An Argument for Incompatilism. Zitiert nach Watson, G. (2003, ed.), *Free Will*. Oxford, New York: Oxford University Press, 38-57
- Weatherford, R.C. (2005) Freedom and Determinism. In Honderich, T. (ed.), *The Oxford Companion to Philosophy*. Oxford, New York: Oxford University Press, 313-315.
- Wolfram, S. (2001) *A New Kind of Science*. Champaign: Wolfram Media.
- Zellulärer Automat(2005, 11.5.) *Wikipedia: The Free Encyclopedia*, http://de.wikipedia.org/wiki/Zellul%C3%A4rer_Automat.
- Zuse, K. (1969) *Rechnender Raum*, Friedrich Vieweg & Sohn.