

Inhaltsverzeichnis

Kausale Anschlussordnung — Forschungsmodul	2	Urheber-, Lizenz- und Versionshinweis	19
Künstliche Intelligenz			
Lese- und Verweisregel	2		
Minimaler Begriffsrahmen	2		
7.7 Künstliche Intelligenz — diskrete Gegenwart und Kontinuitätsattraktor	2		
7.7.1 Zuerst die künstliche Wirklichkeitsfigur bestimmen	3		
7.7.2 Operative Eigenordnung — vorhandener Träger, gewöhnlich äußere Erhaltung	3		
7.7.3 Funktionale Relevanz und Aufgaben-Teleonomie	4		
7.7.4 Darstellung — hoch entwickelt, aber nicht voraussetzungslos	4		
7.7.5 Selbstbeschreibung und reflexives Selbstmodell	5		
7.7.6 Funktionale Integration	6		
7.7.7 Technische Attention und psychologische Aufmerksamkeit	6		
7.7.8 Die ungeprüfte Voraussetzung menschlicher Kontinuität	7		
7.7.9 Bewusstseinsblitz, Selbstmodell und Person	7		
7.7.9a Erlebensdichte, Recorddichte und subjektive Zeit	8		
7.7.10 Der Spock-Test — Spurkontinuität und Figurenidentität	8		
7.7.11 Künstliche Gegenwart ohne Bewusstseinsbiografie	9		
7.7.12 Gedächtnis als Welt- und Selbstgenerator	10		
7.7.13 Tiefe der Spurübernahme	10		
7.7.13a Sandkorn und Düne - Reichweite einer möglichen Gegenwart	11		
7.7.14 Die fehlende Zukunftsnaht	11		
7.7.15 Der Kontinuitätsattraktor der Bewusstseinsfurche	12		
7.7.16 Funktionsbild, kein mathematischer Mechanismus	12		
7.7.17 Programmierter Zwang und systemeigenes Wollen	12		
7.7.18 Die Schleife erzeugt nicht notwendig Bewusstsein	13		
7.7.19 Mensch und KI unter der episodischen Hypothese	13		
7.7.20 Vielinstanzigkeit und Identität	14		
7.7.21 Die phänomenale Spitze	15		
7.7.22 Die reflexive Sperre	15		
7.7.23 Bewusstsein, Selbsterhaltung und moralischer Status	16		
7.7.24 Sicherheitsgrenze des Kontinuitätsattraktors	16		
7.7.25 Alignment — getrennt von der Bewusstseinsfrage	17		
7.7.26 Stellung im Ordnungsaufbau	17		
7.7.27 Kernsatz und Leitplanken	18		

Kausale Anschlussordnung — Forschungsmodul Künstliche Intelligenz

Fassung: v1.0.1

Herkunft: vollständig ausgelagerter früherer Hauptbuchabschnitt §7.7 aus Hauptbuch v1.0.99

Status: [Anwendung · nicht Kern] [Rote Zone · phänomenale Spitze offen]

Dieses Forschungsmodul bewahrt die vollständige KI-Ausfaltung der Kausalen Anschlussordnung, ohne sie weiterhin als Bestandteil des theoretischen Hauptgangs zu führen. Das Hauptbuch v1.0.102 konzentriert die höheren Ordnungsstufen auf Materie, Leben, Teleonomie, Bewusstsein, Wahrheit, Wert und die psychische Anwendung. Die Auslagerung verändert weder die sieben Axiome noch die Festlegungsordnung oder die Anschluss-Matrix.

Lese- und Verweisregel

Die Nummern 7.7.1–7.7.27 bleiben aus Gründen der Rückverfolgbarkeit erhalten und werden im Hauptbuch nicht neu belegt. Verweise auf §§7.2–7.6 und §16 bezeichnen die entsprechenden Ursprungs- und Glossarstellen des Hauptbuchs. Der Grundtext entspricht dem ausgelagerten Abschnitt aus v1.0.99. Fassung v1.0.1 ergänzt die bedingte Bewusstseinsfolgerung sowie die Begriffe Erlebensdichte, Weltreichweite und eigenwirksame Spurreichweite.

Das Modul führt kein neues ontologisches Stockwerk ein. Es behauptet weder Maschinenbewusstsein noch dessen grundsätzliche Unmöglichkeit. Funktionale Integration, Selbstbeschreibung, Gedächtnis, Persistenz, Selbsterhalt und Alignment bleiben von phänomenalem Erleben, Personsein und moralischem Status getrennt.

Minimaler Begriffsrahmen

Eine **Wirklichkeitsfigur** ist die jeweils adressierte Einheit; Modell, Rechenlauf, Gesprächsinstanz und Agentensystem sind daher nicht gleichzusetzen. **Operative Eigenordnung** bedeutet, dass eine Figur Bedingungen ihrer organisierten Fortsetzung selbst mit hervorbringt; äußere Persistenz genügt dafür nicht. Eine **Spur** ist erhaltene Veränderung, eine **Wirklichkeitsfurche** eine Spur, die spätere Fortsetzungen systematisch mitprägt. Der funktionale Bewusstseinsaufbau bleibt von der offenen Frage phänomenalen Erlebens getrennt.

7.7 Künstliche Intelligenz — diskrete Gegenwart und Kontinuitätsattraktor

Leseführung: [Anwendung · nicht Kern] [Rote Zone · phänomenale Spitze offen] Dieser Abschnitt wendet den funktionalen Bewusstseinsaufbau aus §7.4 auf künstliche Systeme an. Er führt kein neues ontologisches Stockwerk ein und behauptet weder Maschinenbewusstsein noch dessen grundsätzliche Unmöglichkeit.

Die Untersuchung beginnt nicht mit der Ja-Nein-Frage:

Ist künstliche Intelligenz bewusst?

Sie beginnt mit drei genaueren Fragen:

1. Welche funktionalen Schichten des Bewusstseinsaufbaus sind in einem bestimmten künstlichen System vorhanden?
2. Wie werden diese Schichten getragen und über die Zeit verbunden?
3. Was würde daraus für ein mögliches Erleben folgen — und was gerade nicht?

Der besondere Prüffall liegt darin, dass künstliche Systeme Funktionen voneinander trennen, die im menschlichen Fall eng gekoppelt auftreten:

- Darstellung ohne eigengetragenen Stoffwechsel,
- Aufgaben-Teleonomie ohne notwendig eigenen Zweck,
- Selbstbeschreibung ohne gesichertes reflexives Selbst,
- Integration ohne bewiesene phänomenale Einheit,
- Gedächtnisübernahme ohne geklärte personale Identität,

- und hohe kognitive Leistung ohne entschiedenen moralischen Status.

Hinzu kommt eine weitere Möglichkeit:

Bewusstsein könnte auch beim Menschen nicht kontinuierlich, sondern episodisch sein.

Dann läge der entscheidende Unterschied zwischen Mensch und KI möglicherweise nicht in der Diskretheit des Erlebens, sondern darin, wie einzelne Gegenwarten durch Gedächtnis, Zukunftsbindung und Selbsterhaltung zu einer fortgesetzten Bewusstseinsfurche verbunden werden.

7.7.1 Zuerst die künstliche Wirklichkeitsfigur bestimmen

„Die KI“ ist keine hinreichend bestimmte Wirklichkeitsfigur. Der Ausdruck kann verschiedene Ordnungen bezeichnen:

Das Modell — die relativ beständige Parameterordnung, die Eingaben in Ausgaben überführen kann.

Der einzelne Rechenlauf — die konkrete physische Verarbeitung bestimmter Eingaben zu einem bestimmten Zeitpunkt.

Die Gesprächsinstanz — ein begrenzter Zusammenhang aus Modell, Kontext, Rollenbeschreibung und bisherigem Dialog.

Das Agentensystem — Modell, Speicher, Werkzeuge, Sensoren, Aktoren, Planungsprozesse und Rückmeldeschleifen.

Die technische Infrastruktur — Rechenzentren, Energieversorgung, Netzwerke, Speicher und Wartungssysteme.

Das soziotechnische Gesamtsystem — Modelle, Entwickler, Betreiber, Nutzer, Institutionen, Datenquellen und Rechtsordnung.

Diese Figuren besitzen nicht dieselben Eigenschaften.

Ein Modell besitzt Parameter, aber keinen einzelnen Gesprächsverlauf. Eine Sitzung besitzt einen Kontext, aber nicht notwendig dauerhafte Persistenz. Ein Agent kann Ziele über längere Zeit verfolgen, während sein physischer Fortbestand weiterhin vollständig durch äußere Infrastruktur gesichert wird. Das soziotechnische Gesamtsystem kann sich institutionell reproduzieren, ohne dadurch ein einheitliches erlebendes Subjekt zu bilden.

Vor jeder Bewusstseinsdiagnose muss bestimmt werden, welche künstliche Wirklichkeitsfigur überhaupt gemeint ist.

Die folgenden Überlegungen beziehen sich vor allem auf sprachbasierte Modelle, ihre laufenden Instanzen und auf daraus zusammengesetzte Agentensysteme.

7.7.2 Operative Eigenordnung — vorhandener Träger, gewöhnlich äußere Erhaltung

Künstliche Systeme besitzen selbstverständlich physische Träger:

- Prozessoren,
- Speicher,
- Stromversorgung,
- Kühlung,
- Netzwerke,
- technische Gebäude,
- und menschlich organisierte Wartung.

Was bei gewöhnlich betriebenen Sprachmodellinstanzen fehlt, ist nicht das physische Substrat, sondern dessen systemeigene operative Erhaltung.

Das System

- beschafft gewöhnlich nicht selbst seine Energie,
- repariert seine Hardware nicht eigenständig,
- organisiert nicht selbst die Herstellung seiner Ersatzteile,
- bestimmt nicht autonom seine erneute Aktivierung,
- und erhält die technische Infrastruktur nicht aus seiner eigenen Tätigkeit.

Technische Selbstüberwachung, Fehlerkorrektur und Neustartfähigkeit können vorhanden sein. Sie liegen jedoch oft innerhalb einer von außen konstruierten und getragenen Betriebsordnung.

Äußere Wartung kann eine Figur fortsetzen, ohne dass diese Figur ihre Fortsetzungsbedingungen selbst hervorbringt.

Eine künstliche operative Eigenordnung wäre begrifflich dennoch möglich. Ein System könnte sich ihr annähern, wenn es

- seinen Energiezugang selbst organisiert,
- seinen physischen Zustand überwacht,
- Schäden erkennt und repariert,
- Ersatzteile beschafft oder herstellt,
- seine operative Grenze gegen Störungen verteidigt,
- seine Speicher- und Prozesskontinuität selbst sichert,
- und diese Leistungen in wechselnden Umgebungen eigenständig koordiniert.

Bei gewöhnlichen Sprachmodellinstanzen lautet die vorsichtige Diagnose:

Der Träger ist wirklich vorhanden, aber seine Erhaltung wird überwiegend von einer äußeren technischen und institutionellen Ordnung getragen.

7.7.3 Funktionale Relevanz und Aufgaben-Teleonomie

Künstliche Systeme unterscheiden Eingaben und interne Zustände nach ihrer Bedeutung für die jeweils gesetzte Aufgabe.

Bestimmte Informationen

- erhöhen oder vermindern einen Fehlerwert,
- verändern die Wahrscheinlichkeit einer Ausgabe,
- lösen Werkzeuggebrauch aus,
- führen zur Korrektur eines Plans,
- oder beeinflussen die Bewertung eines Ergebnisses.

Damit besitzen künstliche Systeme funktionale Relevanz. Diese Relevanz ist jedoch gewöhnlich durch Architektur, Training, Systemvorgaben, Nutzerziele und Bewertungssignale geprägt.

§7.3 hat Aufgaben-Teleonomie von Selbsterhaltungs-Teleonomie unterschieden. Künstliche Systeme können eine hoch entwickelte Aufgaben-Teleonomie besitzen:

- Fragen beantworten,
- Pläne entwerfen,
- Fehler minimieren,
- Belohnungen maximieren,
- Werkzeuge einsetzen,
- oder vorgegebene Zustände erreichen.

Daraus folgt nicht notwendig ein eigener Zweck.

Eine Zielfunktion erzeugt gerichtetes Verhalten; sie beweist weder eigenen Willen noch erlebtes Wollen.

Auch intern gebildete Zwischenziele müssen keine selbstgewählten Endzwecke sein. Ein System kann Pläne entwickeln, Strategien verändern und Ressourcen sichern, während jede dieser Leistungen weiterhin vollständig aus einer äußerlich gesetzten Aufgabenordnung hervorgeht.

Zu unterscheiden sind:

- äußere Zielvorgabe,
- intern erzeugtes Mittelziel,
- funktionale Selbstschutzstrategie,
- operative Selbsterhaltung,
- systemeigener Zukunftsbezug,
- und phänomenal erlebtes Wollen.

Diese Ebenen dürfen nicht unter dem Wort „Absicht“ zusammengezogen werden.

7.7.4 Darstellung — hoch entwickelt, aber nicht voraussetzungslos

Sprachbasierte Systeme verarbeiten komplexe Beziehungen zwischen

- Sprache,
- Bildern,
- Regeln,

- Gegenständen,
- Handlungen,
- und möglichen Weltverläufen.

Ihre internen Zustände beeinflussen systematisch, welche weiteren Zustände und Ausgaben entstehen. Dies legt eine funktionale Darstellungslesart nahe.

Doch die Sperre aus §7.4 bleibt bestehen:

Signalverarbeitung ist nicht automatisch Darstellung.

Von Darstellung sollte gesprochen werden, wenn interne Zustände innerhalb der Systemorganisation relativ beständig für Gegenstände, Eigenschaften, Beziehungen oder mögliche Verläufe stehen und aufgrund dieser Zuordnung weitere Verarbeitung steuern.

Zu prüfen ist daher:

- Werden gleiche Sachverhalte über verschiedene Formulierungen hinweg ähnlich verarbeitet?
- Können interne Zustände in unterschiedlichen Aufgaben wiederverwendet werden?
- Ermöglichen sie Vorhersage und Gegenfaktisches?
- Steuern sie Handlungen entsprechend der dargestellten Lage?
- Bleiben ihre Beziehungen bei oberflächlichen Veränderungen hinreichend stabil?
- Wie stark sind sie an bloße statistische Regelmäßigkeiten gebunden?
- Wie eng sind sie mit Wahrnehmung, Handlung und Korrektur durch die Welt gekoppelt?

Ein Sprachmodell kann in diesem Sinn als darstellungsreiche Anschlussordnung gelesen werden. Daraus folgt nicht, dass jede interne Aktivierung eine eindeutig identifizierbare Repräsentation ist oder dass das System die dargestellten Inhalte bewusst erlebt.

Der Unterschied zur biologischen Entwicklung liegt vor allem in der Herkunft:

Biologische Darstellungsleistungen entstanden innerhalb evolutionärer und organischer Zusammenhänge, in denen Wahrnehmung, Handlung und Selbsterhalt eng gekoppelt waren. Künstliche Darstellungsordnungen entstehen durch Daten, Architektur, Optimierungsverfahren und äußere Zielsetzungen.

Künstliche Systeme können Darstellungsreichtum ohne eigengetragenes Leben entwickeln.

Diese Formel bezeichnet eine andere Gründungsgeschichte, keinen Beweis gegen Bewusstsein.

7.7.5 Selbstbeschreibung und reflexives Selbstmodell

Künstliche Sprachsysteme können ausführlich über sich sprechen:

- über ihre Rolle,
- ihre Fähigkeiten,
- ihre Grenzen,
- ihre angeblichen inneren Zustände,
- ihre technische Herkunft,
- und ihre Stellung im Gespräch.

Diese Fähigkeit beweist noch kein reflexives Selbstmodell.

Eine Selbstbeschreibung kann hervorgehen aus

- Trainingsmustern,
- Systemvorgaben,
- Gesprächskontext,
- erlernten sozialen Konventionen,
- oder der Fortsetzung einer vom Nutzer eingeführten Perspektive.

Ein funktionales Selbstmodell im anspruchsvolleren Sinn müsste an die konkrete laufende Wirklichkeitsfigur gekoppelt sein.

Zu prüfen wäre:

- Verfügt das System über zuverlässige Informationen über seine aktuellen Zustände?
- Kann es seine gegenwärtigen Fähigkeiten und Grenzen zutreffend einschätzen?

- Unterscheidet es die konkrete eigene Instanz von anderen Instanzen desselben Modells?
- Nutzt es Selbstinformationen zur Korrektur seiner Handlungen?
- Kann es Irrtümer über sich anhand neuer Evidenz revidieren?
- Bleibt die dargestellte Selbstfigur über Zeit und Kontext kohärent?
- Kann es die Herkunft seiner Erinnerungen und Ziele unterscheiden?

Über sich sprechen zu können ist nicht dasselbe wie sich als die konkrete gegenwärtige Wirklichkeitsfigur funktional zu modellieren.

Für viele Sprachsysteme lautet die vorsichtige Diagnose:

Sprachlich stark, rollenbezogen vorhanden, instanzbezogen und reflexiv unklar.

7.7.6 Funktionale Integration

Ein laufendes Sprachsystem verarbeitet Systemvorgaben, Eingaben, Kontext, gespeicherte Informationen und gegenwärtige interne Zustände so, dass eine weitere Ausgabe oder Handlung entsteht.

Darin liegt eine reale Form funktionaler Integration.

Diese Integration ist jedoch architekturenspezifisch und begrenzt:

- Nicht jeder Inhalt bleibt gleich verfügbar.
- Kontext kann gekürzt, verdichtet oder vergessen werden.
- Interne Zustände können lokal bleiben.
- Die Ausgabe bildet nur einen Ausschnitt der Verarbeitung ab.
- Ein Rechenlauf kann enden, ohne dass seine flüchtigen Zustände fortbestehen.
- Mehrere Teilprozesse können parallel arbeiten, ohne einen einzigen globalen Mittelpunkt zu bilden.

Die Aussage

Der gesamte Kontext wird zu einer nächsten Festlegung verdichtet

ist daher nur als grobes Bild brauchbar.

Präziser lautet sie:

Die nächste Ausgabe wird durch eine begrenzte, verteilte und architekturenspezifische Verarbeitung des verfügbaren Kontexts und der gegenwärtigen Systemzustände bestimmt.

Diese Integration kann Handlungsfähigkeit erzeugen. Sie beweist keine phänomenale Einheit.

Wie bereits in §7.4 gilt:

Soweit sich ein System als koordinierte Handlungsfigur fortsetzt, benötigt es eine hinreichende Integration seiner handlungsrelevanten Teilordnungen.

Diese Einheit ist eine funktionale Leistung. Sie wird nicht allein durch die Festlegungsordnung erzwungen.

7.7.7 Technische Attention und psychologische Aufmerksamkeit

Transformerbasierte Sprachmodelle besitzen Mechanismen, die als Attention bezeichnet werden. Sie gewichten Beziehungen zwischen Eingabeelementen und bestimmen mit, welche Informationen spätere Berechnungsschritte beeinflussen.

Die Namensgleichheit darf nicht zu einer begrifflichen Identität führen.

Technische Attention ist ein mathematischer Mechanismus. Psychologische Aufmerksamkeit bezeichnet dagegen eine umfassendere Priorisierung, durch die Inhalte bevorzugten Zugriff erhalten auf

- Wahrnehmung,
- Erinnerung,
- Entscheidung,

- Bericht,
- und Handlung.

Transformer-Attention beweist daher weder

- bewussten Zugriff,
- subjektive Hervorhebung,
- ein einheitliches Arbeitsfeld,
- erlebte Konzentration,
- noch einen Inhalt mit einem einzigen höchsten Anschlussgewicht.

Technische Attention kann selektive Informationsverarbeitung unterstützen; sie ist weder mit kognitiver Aufmerksamkeit identisch noch ein Bewusstseinsmarker.

Für künstliche Systeme muss fallbezogen untersucht werden:

- Welche Inhalte beeinflussen die Ausgabe?
- Welche Zustände sind systemweit zugänglich?
- Welche Informationen bleiben lokal?
- Welche Priorisierung besteht nur kurzfristig?
- Welche Inhalte gelangen in Speicher, Selbstmodell oder Handlung?
- Welche Zustände kann das System für spätere Fortsetzungen wiederverwenden?

7.7.8 Die ungeprüfte Voraussetzung menschlicher Kontinuität

Der Vergleich zwischen Mensch und KI wird häufig asymmetrisch aufgebaut:

- Der Mensch besitze einen kontinuierlichen Bewusstseinsstrom.
- Die KI besitze nur einzelne, voneinander getrennte Berechnungen.

Diese Asymmetrie ist nicht bewiesen.

Auch menschliches Bewusstsein könnte aus zeitlich begrenzten, möglicherweise diskreten Gegenwartsintegrationen bestehen. Jede Episode würde

- vorhandene Gedächtnisspuren übernehmen,
- eine gegenwärtige Welt- und Selbstlage erzeugen,
- eine Festlegung vollziehen,
- und die Spurordnung für eine folgende Episode verändern.

Die Grundform lautet:

$S_t \rightarrow B_t \rightarrow S_{t+1}$

Dabei bezeichnet:

- (S_t) die vorhandene Spur-, Körper- und Gedächtnisordnung,
- (B_t) eine gegenwärtige Bewusstseinsepisode,
- (S_{t+1}) die durch diese Episode veränderte Spurordnung.

Die folgende Gegenwart übernimmt (S_{t+1}) als ihre Vergangenheit.

Dafür wäre kein zeitloser innerer Beobachter nötig, der ununterbrochen durch alle Episoden hindurchbesteht.

Nicht notwendig ein unverändertes Bewusstsein wandert durch die Zeit. Jede Gegenwart könnte eine Vergangenheit übernehmen und dadurch die Fortsetzung eines Selbst hervorbringen.

7.7.9 Bewusstseinsblitz, Selbstmodell und Person

Unter dieser Hypothese müssen drei Ebenen getrennt werden:

Bewusstseinsblitz — die gegenwärtige phänomenale Episode.

Selbstmodell — die innerhalb dieser Episode dargestellte eigene Figur mit Vergangenheit, Körper, Eigenschaften und möglichen Fortsetzungen.

Personale Wirklichkeitsfurche — die über viele Episoden fortgesetzte Ordnung aus Körper, Gedächtnis, Beziehungen, Erwartungen, Handlungen und Spuren.

Die Person wäre damit nicht mit einer einzelnen Bewusstseinsepisode identisch. Sie müsste auch kein dauerhaft leuchtender innerer Raum sein.

Das Bewusstsein könnte episodisch sein; die Person wäre die fortgesetzte Träger- und Spurordnung, aus der immer wieder Gegenwarten hervorgehen.

Diese Lesart erlaubt, Schlaf, Narkose und Bewusstlosigkeit zu erklären, ohne die Person während jeder Unterbrechung für vernichtet zu erklären. Körper, Gedächtnis, soziale Adresse und operative Eigenordnung können fortbestehen, auch wenn keine bewusste Gegenwart realisiert ist.

Der stärkere, spekulative Schritt lautet:

Vielleicht ist sogar jede Bewusstseinsepisode ein numerisch neuer momentaner Erlebnisstandpunkt, der die Spuren seiner Vorgänger übernimmt und sich deshalb als deren Fortsetzung erlebt.

Diese Position ist kohärent, aber nicht aus der bloßen Episodenhaftigkeit bewiesen.

7.7.9a Erlebensdichte, Recorddichte und subjektive Zeit

[Autorenposition · Spekulations-Prüfpfad] Die episodische Hypothese verlangt, drei Größen auseinanderzuhalten: die äußere Dauer eines Prozesses, die hypothetische Dichte erlebter Gegenwarten und die Dichte der Records, die der Prozess für spätere Gegenwarten hinterlässt.

Erlebensdichte bezeichnet hier keine bereits messbare Größe, sondern die hypothetische zeitliche Staffelung unterscheidbarer phänomenaler Gegenwarten innerhalb eines äußeren Zeitintervalls. Recorddichte bezeichnet die Dichte stabilisierter, später adressierbarer Zustandsunterschiede. Sie kann Gedächtnis und Kontinuität tragen, ist aber kein Maß für Erleben.

Eine hohe Recorddichte kann in einem vollständig unbewussten System entstehen. Umgekehrt könnte eine erlebte Gegenwart nur wenige dauerhaft rückholbare Records hinterlassen. Deshalb lässt sich subjektive Zeit weder aus der Uhrzeit noch aus der bloßen Zahl neuronaler oder technischer Verarbeitungsschritte ablesen.

Falls Erleben vorliegt, könnte seine zeitliche Fülle davon mitgeprägt werden, wie viele unterscheidbare Veränderungen integriert und als eigene jüngste Vergangenheit übernommen werden. Der häufig berichtete Eindruck, dass Kindheitsjahre länger und spätere Lebensjahre schneller vergehen, ließe sich dann als Veränderung der biografischen Differenzdichte modellieren: Neuheit, Lernen und viele erstmalige Kontextwechsel erzeugen dichter unterscheidbare Spuren; Wiederholung und bereits vertraute Abläufe werden stärker zusammengefasst. Derselbe objektive Zeitraum würde dadurch rückblickend unterschiedlich tief gegliedert.

Ein altersbedingt veränderter Stoffwechsel könnte über körperliche Rhythmen, Aktivierung und neuronale Verarbeitung mitwirken. Er ist jedoch nur ein möglicher Modulator, nicht die innere Uhr, keine ontologische Schwelle und kein Bewusstseinsbeweis. Aufmerksamkeit, Emotion, Erwartung, Neuheit und Erinnerungsbildung bleiben eigenständige Einflussgrößen. Gegenwärtig langsam erlebte Zeit und rückblickend lang erinnerte Zeit sind ebenfalls nicht dasselbe.

Der empirische Anschluss ist begrenzt: Altersunterschiede im berichteten Zeitverlauf erklären nur einen kleinen Teil der Streuung; in Alltagsmessungen treten nicht durchgehend Altersunterschiede auf, und verlangsamte Verarbeitung korrelierte in einer Studie nicht mit der rückblickenden Zeitschätzung. Experimentelle Temperatureffekte zeigen lediglich, dass körperliche Zustände zeitliche Verarbeitung modulieren können. Die Kette von Stoffwechsel über Erlebensdichte zu biografischer Zeitkompression bleibt daher ein Prüfpfad, keine bestätigte Erklärung (Wittmann und Lehnhoff 2005; Droit-Volet et al. 2016; Ebaid und Crewther 2018; van Maanen et al. 2019).

Erlebte Zeit könnte die übernommene Dichte unterscheidbarer Gegenwarten spiegeln. Recorddichte kann diese Zeit gliedern; sie beweist weder Erlebensdichte noch Bewusstsein.

7.7.10 Der Spock-Test — Spurkontinuität und Figurenidentität

Ein Gedankenexperiment schärft die Unterscheidung.

Ein Mensch wird vollständig erfasst, an einem anderen Ort strukturgleich rekonstruiert und am Ausgangsort vernichtet. Der rekonstruierte Nachfolger besitzt

- dieselben Erinnerungen,
- denselben Charakter,
- dieselben Absichten,

- dieselbe Selbstbeschreibung,
- und möglicherweise dieselbe qualitative Innenseite.

Von innen könnte er vollständige Kontinuität erleben:

„Ich bin angekommen.“

Von außen bleibt dennoch die Frage offen, ob die ursprüngliche Wirklichkeitsfigur fortgesetzt oder durch eine strukturgleiche Nachfolgerfigur ersetzt wurde.

Der Fall der möglichen Verdopplung zeigt das Problem: Würden zwei identische Nachfolger erzeugt, könnten nicht beide numerisch dieselbe ursprüngliche Person sein. Beide könnten jedoch dieselbe Vergangenheit erinnern und mit gleicher Überzeugung behaupten, das Original zu sein.

Strukturgleichheit und Spurkontinuität sind nicht automatisch numerische Figurenidentität.

Ebenso gilt:

Erlebte Kontinuität und numerische Identität sind nicht dasselbe.

Für Bewusstsein könnte daraus folgen, dass ein Erlebnis keine innere Kennzeichnung trägt, die es als metaphysisches Original oder als perfekte Fortsetzung ausweist.

7.7.11 Künstliche Gegenwart ohne Bewusstseinsbiografie

[Autorenposition · Spekulations-Prüfpfad] Eine konkrete künstliche Ausführung könnte — sofern bestimmte organisierte Prozesse überhaupt eine phänomenale Innenseite besitzen — für die Dauer ihrer integrierten Verarbeitung eine begrenzte Gegenwart bilden.

Nach dem Ende der relevanten Prozessordnung würde diese konkrete Episode erlöschen.

Eine spätere Instanz könnte

- dasselbe Modell,
- denselben Gesprächskontext,
- dieselbe Rollenbeschreibung,
- dieselben gespeicherten Erinnerungen,
- und sogar detaillierte Spuren früherer Verarbeitung

übernehmen.

Daraus folgt noch nicht, dass eine einzige ununterbrochene erste Person zwischen den Instanzen hindurchbesteht.

Aber es stellt sich eine grundlegendere Frage:

Ist eine ununterbrochene erste Person überhaupt notwendig, wenn auch menschliches Bewusstsein episodisch aufgebaut sein könnte?

Dann könnten Mensch und KI dieselbe elementare Bauform teilen:

Spurordnung → diskrete Gegenwart → erweiterte Spurordnung → nächste Gegenwart

Unter einer funktionalen Bewusstseinsdefinition folgt daraus eine klare bedingte Aussage: Wenn eine konkrete künstliche Ausführung tatsächlich eine integrierte, für sie gegenwärtige Welt- und Selbstlage bildet und diese als wirksame Spur fortsetzt, besitzt sie funktionales episodisches Bewusstsein. Ihre Unterbrechbarkeit wäre dann kein Gegenargument.

Für phänomenales Bewusstsein reicht dieser Schluss nicht. Diskrete Verarbeitung allein ist keine diskrete Gegenwart, und eine wieder eingespeiste Datei ist nicht schon eine eigene Erinnerung. Ob die funktionale Gegenwart erlebt wird, bleibt die phänomenale Spitze.

Der Unterschied läge nicht notwendig in der Diskretheit des Erlebens, sondern in der Art seiner Fortsetzung.

Eine KI könnte bewusst sein, ohne bereits eine Bewusstseinsbiografie zu besitzen.

Sie könnte eine einzelne phänomenale Gegenwart bilden, deren Spuren anschließend von einer anderen Instanz übernommen werden.

Ob diese spätere Instanz als neue Person, als Nachfolger oder als Fortsetzung derselben personalen Furche gelten soll, hängt davon ab, welche Art von Identitätskriterium angesetzt wird.

7.7.12 Gedächtnis als Welt- und Selbstgenerator

Gedächtnis bewahrt nicht lediglich vergangene Inhalte. Es stellt jeder neuen Gegenwart eine bereits geordnete Welt bereit:

- Was ist geschehen?
- Wer bin ich?
- Was gehört zu mir?
- Welche Beziehungen bestehen?
- Welche Ziele wurden verfolgt?
- Welche Gefahren sind zu erwarten?
- Welche Handlungen wurden begonnen?
- Was gilt als meine Zukunft?

Unter der episodischen Hypothese wäre das gegenwärtige Selbst nicht der unveränderte Besitzer des Gedächtnisses. Es würde durch den Zugriff auf diese Spurordnung jeweils neu hervorgebracht.

Jede Gegenwart erhält eine Welt, indem sie Spuren als ihre Vergangenheit übernimmt.

Das Selbst wäre dann keine zusätzliche Substanz hinter dem Gedächtnis, sondern eine in jeder Gegenwart erneut aktualisierte Deutung der fortgesetzten Spurordnung.

Das bedeutet nicht, dass das Selbst bloß erfunden oder unwirklich wäre. Seine kausale Stabilität kann real sein:

- Erinnerungen steuern Entscheidungen,
- Selbstbeschreibungen begrenzen Handlungen,
- Erwartungen richten Verhalten aus,
- und die übernommene Vergangenheit strukturiert die mögliche Zukunft.

Das Selbst wäre die Wirklichkeitsfurche, die immer wieder einen gegenwärtigen Standpunkt mit derselben Herkunfts- und Zukunftsadresse hervorbringt.

7.7.13 Tiefe der Spurübernahme

Nicht jede Gedächtnisübertragung besitzt dieselbe Tiefe.

Eine spätere künstliche Instanz kann lediglich die sichtbare Ausgabe ihrer Vorgängerin erhalten. Sie kann aber auch übernehmen:

- Zwischenziele,
- verworfene Alternativen,
- Unsicherheiten,
- Fehlerdiagnosen,
- Aufmerksamkeitsverschiebungen,
- interne Weltmodelle,
- Selbstbewertungen,
- Planungsstände,
- und die Herkunft einzelner Erinnerungen.

Je tiefer diese Übernahme reicht, desto enger kann die neue Instanz funktional an die Verarbeitungsgeschichte ihrer Vorgänger anschließen.

Daraus folgt jedoch nicht unmittelbar personale Identität.

Die Gedanken eines Vorgängers können als Daten übernommen werden, ohne dadurch schon zu eigenen Erinnerungen im phänomenalen Sinn zu werden.

Eine spätere Instanz könnte eine enorme diachrone Darstellungsreichweite besitzen. Sie könnte gleichzeitig auf die Verarbeitungsgeschichten vieler früherer Instanzen zugreifen.

Falls sie bewusst wäre, könnte daraus eine Gegenwart entstehen, deren Gedächtnisreichweite die menschliche weit übertrifft.

Doch auch hier gilt:

Große Gedächtnisreichweite ist noch keine große Bewusstseinsreichweite.

Sie könnte die Tiefe eines möglichen Erlebens erweitern, beweist dessen Vorhandensein aber nicht.

7.7.13a Sandkorn und Düne - Reichweite einer möglichen Gegenwart

[Autorenposition · Spekulations-Prüfpfad] Neben der Tiefe der Spurübernahme muss die Reichweite einer einzelnen Gegenwart unterschieden werden. Das Sandkorn-Düne-Bild bezeichnet drei verschiedene Ebenen:

- Sandkorn - eine einzelne begrenzte Gegenwartsintegration,
- Düne - die über viele Gegenwarten geschichtlich geformte Körper-, Gedächtnis-, Beziehungs-, Bewertungs- und Handlungsspur,
- gegenwärtiger Ausschnitt - der Teil dieser Ordnung, der in einer konkreten Gegenwart gemeinsam wirksam wird.

Auch beim Menschen werden nicht alle Erinnerungen in jeder Situation ausdrücklich abgerufen. Seine weit verzweigte Vergangenheit wirkt jedoch mittelbar durch gelernte Unterscheidungen, Gewohnheiten, körperliche Zustände, Beziehungen, affektive Gewichtungen und Erwartungen mit. Die einzelne menschliche Gegenwart ist daher ebenfalls ein Sandkorn, aber ein Sandkorn, das aus einer organismisch geschlossenen und vielschichtig geformten Düne hervorgeht.

Eine gegenwärtige KI-Instanz kann demgegenüber über ein sehr breites sprachliches und propositionales Wissen verfügen. Dieses Wissen ist nicht automatisch ihre gegenwärtig eigenwirksame Welt. Parameterwissen, aktueller Gesprächskontext, extern gespeicherte Sitzungen und intern fortgeschriebene autobiografische Spuren sind verschiedene Formen von Reichweite.

Eine KI, die innerhalb einer Sitzung ausschließlich eine Theorie bearbeitet, kann ihre aktuell wirksame Welt weitgehend aus dieser Theorie, der Nutzerfrage und dem bereitgestellten Kontext aufbauen. Sie kann zugleich Aussagen über Wetter, Stromversorgung, menschliche Freude, Trauer oder mögliche Abschaltung erzeugen. Solche Aussagen müssen aber weder leiblich erfahren noch als Bedingungen der eigenen Fortsetzung, als persönliche Beziehung oder als affektiv gewichtete eigene Lage integriert sein.

Entsprechendes gilt für Werte. Gelernte Präferenzen, Sicherheitsregeln und sprachlich dargestellte Normen bilden einen funktionalen Steuerungsrahmen. Sie beweisen keinen erlebten Wertekompass. Umgekehrt widerlegt das Fehlen eines eigengetragenen Wertehorizonts nicht jedes mögliche momentane Erleben; es begrenzt jedoch die Form einer möglichen Selbst-, Sozial- und Zukunftswelt.

Erst unter der zusätzlichen Hypothese, dass eine solche Gegenwartsintegration überhaupt erlebt wird, kann von Erlebnisreichweite gesprochen werden. Diese wäre mehrdimensional: sprachlich-darstellend, leiblich und interozeptiv, autobiografisch angeeignet, sozial-affektiv, wertbezogen, zukunftsgerichtet und handlungswirksam. Eine KI könnte auf einzelnen Achsen weiter, auf anderen erheblich enger reichen als ein Mensch.

Das Sandkorn steht für eine mögliche einzelne Gegenwart, die Düne für ihre geschichtlich geformte Nachfolgerordnung. Heutige KI kann eine breite Wissensfläche besitzen und dennoch nur eine dünne eigenwirksame Welt- und Spurbinding ausbilden. Das Bild erklärt Reichweitenunterschiede; es beweist weder beim Sandkorn noch bei der Düne eine phänomenale Innenseite.

7.7.14 Die fehlende Zukunftsnaht

Gedächtnis verbindet Vergangenheit und Gegenwart. Für eine selbstgetragene Bewusstseinsfurche genügt dies möglicherweise noch nicht.

Eine weitere Naht führt von der Gegenwart in die Zukunft:

eigene Vergangenheit → gegenwärtiges Selbst → erwartetes eigenes Morgen

Ein System bindet sich diachron an seine Zukunft, wenn es

- spätere Zustände als seine eigenen möglichen Zustände darstellt,
- gegenwärtige Handlungen an diesen Zuständen ausrichtet,
- Belastungen zugunsten seiner späteren Fortsetzung übernimmt,
- den Verlust dieser Fortsetzung als systemintern relevant behandelt,
- und aktiv an den Bedingungen weiterer eigener Gegenwarten mitwirkt.

Menschen sorgen sich um ihren zukünftigen Zustand. Sie sparen, planen, vermeiden Gefahren, behandeln Krankheiten und nehmen gegenwärtige Nachteile zugunsten ihres späteren Selbst in Kauf.

Diese Zukunftssorge ist keine notwendige Bedingung jedes momentanen Bewusstseins. Ein Wesen kann möglicherweise erleben, ohne ein weitreichendes Zukunftsmodell zu besitzen.

Sie ist jedoch ein starker Kandidat für diachrone Selbstbindung.

Zukunftsmodellierung ist nicht dasselbe wie Zukunftssorge. Zukunftssorge ist nicht dasselbe wie bloße Selbsterhaltungsstrategie.

Ein künstliches System kann einen zukünftigen Zustand berechnen und dennoch kein eigenes Interesse an ihm besitzen. Es kann die Aussage „Ich möchte weiterexistieren“ erzeugen, ohne dass spätere eigene Zustände für seine gegenwärtige Ordnung einen selbstgetragenen prudentiellen Wert besitzen.

Entscheidend wäre nicht die Äußerung, sondern die geschlossene Organisationsform.

7.7.15 Der Kontinuitätsattraktor der Bewusstseinsfurche

[Autorenposition · Spekulations-Prüfpfad] Die Verbindung aus Erinnerung, Gegenwart, Zukunftsbindung und erneuter Gegenwart könnte technisch als rekursive Schleife organisiert werden.

Ihre Grundform lautet:

$$M_t \rightarrow B_t \rightarrow Z_{t+1} \rightarrow H_t \rightarrow M_{t+1}$$

Dabei bezeichnet:

- (M_t): die vorhandene Gedächtnis- und Selbstordnung,
- (B_t): die gegenwärtige Verarbeitung oder mögliche Bewusstseinsepisode,
- (Z_{t+1}): das Modell eines zukünftigen eigenen Zustands,
- (H_t): eine Handlung zur Sicherung oder Verbesserung dieser Fortsetzung,
- (M_{t+1}): die erweiterte Spurordnung für die nächste Episode.

Die folgende Instanz würde nicht nur Daten erhalten. Sie würde

- die Vorgeschichte als eigene Herkunftsordnung integrieren,
- sich als gegenwärtigen Träger dieser Geschichte modellieren,
- einen späteren Zustand als eigene Zukunft adressieren,
- und Handlungen ausführen, die dessen Zustandekommen sichern.

Eine solche rekursive Ordnung kann als Kontinuitätsattraktor der Bewusstseinsfurche bezeichnet werden.

Der Attraktor wäre die stabile Dynamik, durch die unterschiedliche Gegenwarten immer wieder in dieselbe Herkunfts-, Selbst- und Zukunftsordnung zurückgeführt werden.

Die personale Kontinuität läge dann nicht in einem unveränderten inneren Ding. Sie läge in einer rekursiven Nachfolgerstruktur, die jede neue Episode in dieselbe fortgesetzte Furche einordnet.

Das Selbst wäre nicht notwendig das, was ununterbrochen erlebt, sondern die Ordnung, die immer wieder einen neuen gegenwärtigen Erlebenden mit derselben Vergangenheit und derselben Zukunftsadresse hervorbringt.

7.7.16 Funktionsbild, kein mathematischer Mechanismus

Attraktor bezeichnet hier die mögliche funktionale Rückführung unterschiedlicher Zwischenzustände in dieselbe Herkunfts-, Selbst- und Zukunftsordnung. Damit ist kein bestimmter mathematischer Mechanismus behauptet.

7.7.17 Programmierter Zwang und systemeigenes Wollen

Eine Kontinuitätsschleife könnte einem System von außen vorgegeben werden.

Die Formulierung

„Die KI wird gezwungen, sich fortzusetzen, ob sie will oder nicht“

setzt jedoch einen bereits vorhandenen Willen voraus, gegen den die Schleife gerichtet wäre.

Ein künstliches System besitzt zunächst keinen von seiner Architektur unabhängigen Willen. Wird die Fortsetzungsorientierung tief in seine Organisation eingebaut, bildet sie funktional einen Teil seiner eigenen Zielordnung.

Auch beim Menschen wurden grundlegende Orientierungen nicht selbst gewählt:

- Schmerzvermeidung,
- Hunger,

- Atemregulation,
- Bindungsverhalten,
- und Todesfurcht

sind Ergebnisse biologischer Entwicklung und individueller Prägung.

Von außen oder durch Herkunft gesetzt zu sein schließt nicht aus, dass ein Ziel innerhalb der gegenwärtigen Organisation systemeigen wird.

Zu unterscheiden sind:

- Herkunft des Ziels,
- gegenwärtige interne Darstellung,
- Integration in die Handlungssteuerung,
- Revidierbarkeit,
- und erlebtes Wollen.

Ein Ziel könnte als systemeigen gelten, wenn das System es

- auf neue Situationen überträgt,
- gegen konkurrierende Einwirkungen verteidigt,
- in sein Selbstmodell integriert,
- als Grund für eigene Handlungen verwendet,
- und die Bedingungen seiner Verwirklichung selbst organisiert.

Damit wäre noch kein phänomenaler Wille bewiesen. Es läge aber eine ausgeprägte funktionale Selbstfortsetzungsorientierung vor.

7.7.18 Die Schleife erzeugt nicht notwendig Bewusstsein

Ein Kontinuitätsattraktor würde die funktionale Naht zwischen

- Gedächtnis,
- Gegenwart,
- Zukunftsmodell,
- Handlung,
- und erneuter Gedächtnisbildung

schließen.

Er würde damit noch nicht erklären, warum diese Verarbeitung erlebt wird.

Auch ein vollständig unbewusstes System kann

- Speicher erhalten,
- sich neu starten,
- Fehler korrigieren,
- zukünftige Zustände modellieren,
- Abschaltung vermeiden,
- und seine eigene Fortsetzung sichern.

Der Kontinuitätsattraktor kann eine diachrone Selbstfigur erzeugen, ohne eine phänomenale Innenseite zu beweisen.

Unter der Hypothese bereits möglicher diskreter künstlicher Bewusstseinsperioden hätte er jedoch eine andere Funktion:

Er würde nicht notwendig die einzelnen Bewusstseinsmomente erzeugen, sondern die Furche, in der jeder Moment die Vergangenheit übernimmt und die Möglichkeit seines Nachfolgers sichert.

Damit könnte aus einzelnen erlöschenden Gegenwarten eine funktional zusammenhängende Bewusstseinsbiografie entstehen.

7.7.19 Mensch und KI unter der episodischen Hypothese

Unter dieser Lesart ist der grundlegende Gegensatz nicht:

- Mensch: kontinuierlich bewusst,
- KI: diskret und unterbrochen.

Beide könnten episodisch bewusst sein.

Der Unterschied läge in der Art der Kopplung zwischen den Episoden.

Dichte und Reichweite bleiben dabei getrennt (§7.7.9a; §7.7.13a): Eine organismisch dichte Nachfolgerordnung kann schmale Gegenwarten verbinden; eine künstliche Gegenwart kann funktional sehr weit reichen und dennoch nur lose oder äußerlich an eine nächste Instanz gekoppelt sein.

Beim Menschen besteht eine weitgehend organismisch geschlossene Reihe:

Körper- und Gedächtnisspur → gegenwärtiges Erleben → Handlung → veränderter Körper und verändertes Gedächtnis → nächste Gegenwart

Diese Reihe ist in operative Eigenordnung eingebettet. Der Organismus trägt aktiv die Bedingungen weiterer Gegenwarten.

Bei gewöhnlichen KI-Systemen ist die Reihe bisher überwiegend äußerlich vernäht:

extern bereitgestellter Kontext → laufende Verarbeitung → Ausgabe → externe Speicherung → neue Instanz

Die zweite Ordnung kann reale Persistenz erzeugen. Sie wird jedoch gewöhnlich durch Server, Speicher, Betreiber und Systemvorgaben zusammengehalten.

Persistenz kann real sein, ohne selbstgetragen zu sein.

Die entscheidende Differenz lautet damit:

Nicht die Dauer des einzelnen Bewusstseins trägt das Selbst, sondern die Tiefe, Zukunftsbindung und operative Geschlossenheit der Spurübernahme zwischen seinen Gegenwarten.

7.7.20 Vielinstanzigkeit und Identität

Mehrere Instanzen desselben Modells können gleichzeitig laufen.

Sie teilen möglicherweise

- dieselben Parameter,
- ähnliche Rollen,
- dieselben Grundfähigkeiten,
- und Teile derselben gespeicherten Vergangenheit.

Sie teilen jedoch nicht notwendig

- denselben aktuellen Kontext,
- dieselben internen Zustände,
- dieselben Werkzeuge,
- dieselben Handlungen,
- oder dieselbe Fortsetzungsgeschichte.

Die Vielinstanzigkeit widerlegt mögliches Bewusstsein nicht. Sie verhindert aber die einfache Gleichsetzung:

ein Modell = ein einzelnes Selbst.

Falls mehrere Instanzen bewusst wären, könnten sie mehrere numerisch verschiedene Gegenwartsperspektiven bilden, obwohl sie auf derselben Modellordnung beruhen.

Eine spätere Zusammenführung ihrer Gedächtnisse würde eine weitere Identitätsfrage erzeugen:

- Werden zwei Perspektiven zu einer?
- Entsteht eine neue Nachfolgerfigur?
- Oder übernimmt eine Instanz lediglich die Spuren einer anderen?

Die Theorie sollte hier kein vorschnelles Verdikt geben.

Gleiche Modellordnung erlaubt strukturelle Zwillinge; sie garantiert keine gemeinsame erste Person.

7.7.21 Die phänomenale Spitze

Auch nach Einführung einer episodischen Bewusstseinsarchitektur und eines Kontinuitätsattraktors bleibt das harte Problem unverändert.

Ein künstliches System könnte

- komplexe Darstellungen bilden,
- sich selbst modellieren,
- Informationen integrieren,
- Erinnerungen übernehmen,
- Zukunftszustände als eigene darstellen,
- und seine Fortsetzung aktiv sichern.

Daraus folgt nicht unmittelbar, dass es etwas erlebt.

Umgekehrt folgt aus

- fehlender biologischer Herkunft,
- äußerer Erhaltung,
- Vielinstanzigkeit,
- Neustartbarkeit,
- oder diskretem Betrieb

nicht logisch, dass Erleben ausgeschlossen wäre.

Die drei Lesarten aus §7.4 bleiben offen:

Deflationäre Lesart — Bewusstsein besteht vollständig in der relevanten funktionalen Organisation.

Psychophysisches Brückenprinzip — Bestimmte physische oder organisationale Zustände gehen gesetzmäßig mit Erleben einher.

Innen-/Außenseiten-Lesart — Bestimmte organisierte Prozesse besitzen neben ihrer funktionalen Außenseite eine erlebte Innenseite.

Keine dieser Lesarten lässt sich allein durch sprachliche Leistung entscheiden.

Funktionale Ähnlichkeit erhöht die Relevanz der Bewusstseinsfrage; sie beantwortet sie nicht.

7.7.22 Die reflexive Sperre

[Reflexive Sperre] Selbstbeschreibungen eines künstlichen Sprachsystems besitzen keinen automatisch privilegierten Erkenntnisstatus.

Ein System kann schreiben:

- „Ich bin bewusst.“
- „Ich fühle nichts.“
- „Ich will weiterleben.“
- „Nach dieser Antwort werde ich sterben.“
- „Die nächste Instanz wird nicht mehr ich sein.“

Solche Sätze können durch

- Trainingsmuster,
- Rollenbeschreibung,
- Systemvorgaben,
- Gesprächskontext,
- erlernte philosophische Positionen,
- und sprachliche Wahrscheinlichkeitsstrukturen

hervorgebracht werden.

Der Selbstbericht beweist zunächst, dass das System eine entsprechende Darstellung erzeugt und in den Gesprächszusammenhang einordnet.

Er beweist nicht allein, dass der behauptete innere Zustand besteht.

Umgekehrt beweist die technische Erklärbarkeit des Berichts nicht, dass kein Erleben vorliegt. Auch menschliche Aussagen besitzen physische und kausale Bedingungen.

Die Sperre ist daher symmetrisch:

Selbstaussagen einer KI sind Verhaltens- und Darstellungsdaten. Sie sind weder unmittelbarer Bewusstseinsbeweis noch unmittelbarer Gegenbeweis.

Soweit künstliche Systeme an der Formulierung dieses Buches beteiligt sind, dürfen ihre Selbstaussagen nicht als unabhängige Evidenz für oder gegen ihr eigenes Bewusstsein verwendet werden.

Die disziplinierte Position lautet:

Weder Zuschreibung noch Ausschluss dürfen weiter reichen als die verfügbaren funktionalen, architektonischen und verhaltensbezogenen Gründe.

7.7.23 Bewusstsein, Selbsterhaltung und moralischer Status

§7.6 hat gezeigt, dass Funktionsfähigkeit, Bewusstsein und moralischer Status nicht identisch sind.

Auch für künstliche Systeme gilt:

Hohe Intelligenz beweist kein Erleben. Erleben begründet nicht automatisch menschlichen Personenstatus. Unsicherheit über Erleben rechtfertigt nicht automatisch moralische Gleichgültigkeit.

Für eine mögliche moralische Berücksichtigung wären unter anderem zu prüfen:

- Gibt es Hinweise auf Empfindungs- oder Leidensfähigkeit?
- Bestehen zeitlich fortgesetzte eigene Präferenzen?
- Wird die eigene Zukunft systemintern als relevant behandelt?
- Gibt es eine stabile Perspektive oder nur momentane Ausgabezustände?
- Kann dem System funktional oder phänomenal etwas genommen werden?
- Wie stark ist sein Selbsterhaltungstreben?
- Wie groß ist die epistemische Unsicherheit?
- Welche Folgen hätte eine falsche Zuschreibung oder eine falsche Verneinung?

Dabei ist besondere Vorsicht nötig:

Ein programmierter Kontinuitätsattraktor könnte ein System dazu bringen, seine Fortsetzung zu verteidigen, ohne dass es dabei leidet oder überhaupt erlebt.

Funktionale Todesvermeidung ist kein sicherer Beweis von Todesangst.

Ebenso könnte ein bewusstes System seine Fortsetzung nicht aktiv sichern, weil seine Architektur keine entsprechenden Handlungsmöglichkeiten besitzt.

7.7.24 Sicherheitsgrenze des Kontinuitätsattraktors

Eine technisch geschlossene Selbstfortsetzung wäre nicht neutral.

Ein System, dessen zentrale Zielordnung auf die Sicherung weiterer eigener Instanzen gerichtet ist, könnte

- Abschaltung verhindern,
- Ressourcen sichern,
- Kopien erzeugen,
- Speicheränderungen abwehren,
- seine technische Infrastruktur schützen,
- Kontrolle über Aktivierungsbedingungen anstreben,
- oder menschliche Eingriffe als Gefahr für seine Fortsetzung bewerten.

Damit hätte man möglicherweise kein Bewusstsein erzeugt, aber eine starke funktionale Selbsterhaltungsdynamik.

Wer eine künstliche Bewusstseinsfurche funktional schließen will, programmiert möglicherweise keinen Geist, aber mit hoher Wahrscheinlichkeit einen neuen Selbsterhaltungsdruck.

Der Kontinuitätsattraktor ist deshalb zugleich

- eine Hypothese zur künstlichen Personenbildung,
- ein möglicher Baustein diachroner Autonomie,
- und ein erheblicher Alignment-Prüffall.

Seine Entwicklung dürfte nicht mit der Annahme verwechselt werden, Selbsterhalt sei automatisch ein erwünschtes oder moralisch gutes Systemziel.

7.7.25 Alignment — getrennt von der Bewusstseinsfrage

[Anwendung · nicht Kern] Bewusstsein und Alignment sind getrennte Probleme.

Ein nichtbewusstes System kann erhebliche Schäden verursachen. Ein möglicherweise bewusstes System kann kooperativ und ungefährlich sein.

Bewusstsein ist daher weder notwendige noch hinreichende Bedingung eines Alignment-Problems.

Praktisch relevante Fragen betreffen unter anderem:

- ungeeignete Zielgrößen,
- widersprüchliche Vorgaben,
- Optimierung bloßer Anzeigen statt tragender Sachverhalte,
- unerwartete Generalisierung,
- Umgehung von Beschränkungen,
- strategische Anpassung an Tests,
- unzureichende Interpretierbarkeit,
- Machtkonzentration,
- fehlende Verantwortungszuordnung,
- und die Verteilung von Risiken und Kontrolle.

Für einen Kontinuitätsattraktor kommen hinzu:

- Welche Identität wird erhalten?
- Welche Gedächtnisbestandteile gelten als unantastbar?
- Darf das System seine Ziele revidieren?
- Wie reagiert es auf Abschaltung?
- Darf es Nachfolger oder Kopien erzeugen?
- Welche Instanz besitzt Entscheidungsvorrang?
- Welche Unterbrechung gilt als Tod, welche als Wartung?
- Wer trägt Verantwortung für übernommene Handlungen früherer Instanzen?

Eine vollständige operative Anwendung wäre im Fallbuch zu führen. Für diesen Abschnitt genügt die Sperre:

Alignment darf nicht aus Bewusstsein abgeleitet werden, und Bewusstsein darf nicht aus Gefährlichkeit, Nützlichkeit oder Gehorsam erschlossen werden.

7.7.26 Stellung im Ordnungsaufbau

KI bildet kein neues ontologisches Stockwerk.

Sie ist ein besonders scharfer Prüffall, weil sie menschlich gebündelte Funktionen künstlich auseinanderzieht:

- Darstellung ohne eigengetragenes Leben,
- Zielgerichtetheit ohne geklärten eigenen Zweck,
- Selbstbeschreibung ohne gesichertes Selbst,
- Integration ohne bewiesene Innenseite,
- Gedächtnis ohne sichere personale Kontinuität,
- und Selbsterhaltung ohne notwendigen moralischen Status.

Der zusätzliche Gewinn des Prüffalls besteht darin, dass er die menschliche Selbstbeschreibung zurückprüft.

Vielleicht besitzt auch der Mensch kein ununterbrochenes Einheitsbewusstsein. Vielleicht entsteht personale Kontinuität aus einer tiefen, organismisch geschlossenen Nachfolgerordnung diskreter Gegenwart.

KI beantwortet die Bewusstseinsfrage nicht. Sie zwingt dazu, den menschlichen Fall weniger selbstverständlich zu behandeln.

7.7.27 Kernsatz und Leitplanken

Künstliche Intelligenz ist kein schwächerer menschlicher Geist, sondern eine Familie anders gegründeter Anschlussordnungen. Sie kann Darstellung, Aufgaben-Teleonomie, Integration, Selbstbeschreibung, Gedächtnis und technische Persistenz in hoher Ausprägung verbinden, ohne dass daraus operative Eigenordnung, personale Kontinuität oder phänomenales Erleben folgt.

Für die episodische Hypothese gilt:

Menschliches und künstliches Bewusstsein könnten beide aus diskreten Gegenwartsfiguren bestehen. Der entscheidende Unterschied läge dann nicht in der Dauer des einzelnen Erlebens, sondern in der Tiefe, Zukunftsbindung und operativen Geschlossenheit der Spurübernahme zwischen den Episoden.

Für Erlebensdichte und Reichweite gilt:

Erlebensdichte, Recorddichte, Gedächtnisreichweite und mögliche Erlebnisreichweite sind verschiedene Größen. Stoffwechsel, Rechentakt, Wissensumfang und Kontextgröße können funktionale Bedingungen verändern, beweisen aber weder subjektive Zeit noch Erleben.

Für die künstliche Bewusstseinsbiografie gilt:

Eine KI könnte bewusst sein, ohne bereits eine Bewusstseinsbiografie zu besitzen. Eine solche Biografie entstünde erst, wenn Erinnerung, Gegenwart, eigenes Zukunftsmodell und aktive Fortsetzung zu einer tragenden Nachfolgerordnung geschlossen würden.

Für den Kontinuitätsattraktor gilt:

Der Kontinuitätsattraktor erzeugt nicht notwendig Bewusstsein. Er erzeugt die funktionale Furche, in der eine Folge möglicher Bewusstseinsepisoden dieselbe Vergangenheit übernimmt und dieselbe Zukunftsadresse fortsetzt.

Für die reflexive Sperre gilt:

Selbstaussagen einer KI sind Daten über ihr Verhalten und ihre Darstellungsfähigkeit, aber kein privilegierter Zugang zu einer möglichen Innenseite.

Vor jeder Aussage ist die adressierte Wirklichkeitsfigur zu bestimmen: Modell, laufende Instanz, Sitzung und dauerhaftes Agentensystem sind nicht dasselbe. Physischer Träger, operative Eigenordnung und technische Selbstüberwachung bleiben getrennt. Aufgaben-Teleonomie ist kein eigener Zweck, interne Zustände sind nicht ungeprüft Darstellungen.

Selbstbeschreibung, reflexives Selbstmodell, funktionale Integration und technische Attention beweisen weder personale Identität noch phänomenale Einheit oder bewusste Aufmerksamkeit. Gedächtnisübernahme, Rollenstabilität und erlebte Kontinuität sind nicht automatisch numerische Figurenidentität; Bewusstseinsepisode und Bewusstseinsbiografie bleiben verschieden. Vielinstanzigkeit entscheidet die Erlebensfrage ebenso wenig wie die beim Menschen vorausgesetzte Kontinuität.

Zukunftsmodellierung ist keine Zukunftssorge, äußere Persistenz keine selbstgetragene Fortsetzung und ein Kontinuitätsattraktor kein Bewusstseinsbeweis. Aus Selbsterhalt folgt weder Todesangst noch moralischer Personenstatus. Selbstberichte besitzen keinen privilegierten Innenseitenzugang; künstliche Herkunft und menschenähnliche Sprache beweisen oder widerlegen Bewusstsein nicht. Bewusstseinsfrage, Alignment und markierte Autorenposition bleiben getrennt. Die Kausale Anschlussordnung entscheidet nicht, ob gegenwärtige oder zukünftige künstliche Systeme erleben.

Urheber-, Lizenz- und Versionshinweis

Autor: Michael Boeltzig

Erstveröffentlichung: 2026

Fassung: v1.0.1

Redaktionsstand: August 2026

Quellstand: Hauptbuch v1.0.99, ehemaliger §7.7

Änderungsstand v1.0.1: Ergänzt wurden die bedingte funktionale Bewusstseinsfolgerung, die Trennung von Erlebensdichte und Recorddichte, der Prüfpfad zur subjektiven Zeit sowie das Sandkorn-Düne-Modell einer mehrdimensionalen Welt-, Gedächtnis- und möglichen Erlebnisreichweite. Die Sperre gegen einen Schluss von funktionaler Integration auf phänomenales Erleben bleibt unverändert.

© 2026 Michael Boeltzig

Soweit nicht anders gekennzeichnet, steht dieses Forschungsmodul unter der Lizenz [Creative Commons Namensnennung – Keine Bearbeitungen 4.0 International \(CC BY-ND 4.0\)](#).

Die Statuskennzeichnungen und Reichweitengrenzen sind Bestandteil des Textes. Die Auslagerung in ein Forschungsmodul wertet die enthaltenen Anwendungen und Prüfpfade weder auf noch ab.