

# Inhaltsverzeichnis

<b>Hypothesentag-Gutachten</b>	<b>1</b>
Die Gewinnerthese . . . . .	1
Initiale Fassung . . . . .	1
2. Die drei initialen Hypothesen . . . . .	1
Reformulierte Fassung (nach Kritischem Professor) . . . . .	2
4. Reformulierung nach Kritischem Professor . . . . .	2
<b>Expertenrunden und Synthese</b>	<b>2</b>
7. Expertenrunde 1 — Unabhängige Gutachten . . . . .	2
8. Expertenrunde 2 — Repliken-Runde . . . . .	4
9. Synthese — Sokrates . . . . .	6
10. Empirie-Brücke (Phase 3.5) . . . . .	7
11. Externe Begutachtung (Phase 4 — Zweitmeinung) . . . . .	8

## Hypothesentag-Gutachten

### Die Gewinnerthese

**Dekoder-Autonomie als Schwelle der geteilten Hierarchietiefe — der Ort der Schwelle ist die funktionale Überlappung, nicht die absolute Modelltiefe**

### Initiale Fassung

#### 2. Die drei initialen Hypothesen

##### H1 (warm) — Dekoder-Autonomie als Schwelle der geteilten Hierarchietiefe

**Kernsatz.** Dekoder-Autonomie ist keine Eigenschaft des Produzenten, des Artefakts oder des Rezipienten für sich, sondern eine Schwellenbedingung an die *Tiefe des geteilten* hierarchischen generativen Modells: Sie tritt ein, sobald Produzent und Rezipient mindestens eine Hierarchieebene *oberhalb* der artefakt-codierenden Ebene teilen. Diese geteilte obere Ebene übernimmt die Funktion, die sonst der direkte interpersonale Kontakt erfüllt — sie liefert die Priors, unter denen die Oberflächenmerkmale des Artefakts eindeutig disambiguiert werden.

**Begründung.** Der `warm_pick` aus [[06 Hypothesentag/2026-05-24]] hatte Dekoder-Autonomie als Schwellenbedingung zwischen Vor- und Oberstufen symbolischer Repräsentation gesetzt und gefragt, ob es eine identifizierbare Modell-Hierarchietiefe gibt, ab der ein generatives Modell sie ermöglicht. Die hier vorgeschlagene Antwort verschiebt den Ort der Schwelle: Sie liegt nicht in der absoluten Tiefe eines einzelnen Modells, sondern in der *Überlappungstiefe* zweier Modelle. In Fristons Architektur heißt das: Ein Bild, eine Schrift, ein Notat wird genau dann ohne Anwesenheit des Urhebers lesbar, wenn die zur Disambiguierung nötige Ebene nicht im Artefakt steht, sondern als kulturell geteilter Prior beim Empfänger schon vorliegt. Cassirers symbolische Form ist dann der Name für genau diese geteilte obere Ebene.

**Falsifikationsbedingung.** Widerlegt, wenn (a) computationale Agenten zuverlässige Dekoder-Autonomie erreichen, deren generative Modelle *nur* die oberflächen-codierende Ebene teilen; oder (b) autonome Dekodierung mit der Tiefe des *Empfänger*-Modells allein skaliert, unabhängig von der Überlappung mit dem Produzenten.

**Quelle.** `warm_pick`.

---

##### H2 (cold) — Die Differenz der Kunst liegt in der Umkehrung der Zeichenrichtung

**Kernsatz.** Was die symbolische Funktion des Kunstwerks von Sprache, Mythos und Wissenschaft unterscheidet, ist eine Umkehrung der Zeichenrichtung: In den übrigen symbolischen Formen ist der Zeichenträger *ersetzbar*, solange die Bedeutung erhalten bleibt; im Kunstwerk ist der Zeichenträger *unersetzbar*, weil seine Bedeutung gerade im Gebundensein an genau diese sinnliche Gestalt besteht.

**Falsifikationsbedingung.** Widerlegt, wenn (a) Kunstwerke existieren, deren Zeichenträger verlustfrei substituierbar ist und die Kunst bleiben; oder (b) nicht-künstlerische symbolische Formen regelmäßig dieselbe Unersetzbarkeit zeigen.

Quelle. cold\_pick.

---

### H3 (frei) — Tradition transportiert Probleme, nicht Lösungen

**Kernsatz.** Was eine Kultur über Generationen tatsächlich weitergibt, ist nicht primär ein Bestand gespeicherter Lösungen, sondern ein Kanon präzise gestellter, ungelöster Probleme: Die Tradition transportiert in ihrer fruchtbaren Schicht Aporien, nicht Antworten.

**Falsifikationsbedingung.** Widerlegt, wenn unter Überlieferungsträgern mit vergleichbarer Anfangsverbreitung die lösungs-geschlossenen im Mittel ebenso lange oder länger produktiv zirkulieren wie die problem-offenen.

Quelle. frei.

---

## Reformulierte Fassung (nach Kritischem Professor)

### 4. Reformulierung nach Kritischem Professor

#### H1 — überarbeitet (warm)

**Kernsatz.** Dekoder-Autonomie tritt ein, sobald Produzent und Rezipient eine generative Ebene teilen, die *abstrakter* ist als die im Artefakt physisch codierte — und diese geteilte Ebene lässt sich unabhängig vom Dekodier-Erfolg über drei Marker messen: Generalisierung auf ungesehene Artefakt-Varianten, Robustheit gegen Oberflächenrauschen und Transfer auf andere Artefakte derselben Klasse.

**Was sich ändert.** Gegen den Zirkelvorwurf wird die geteilte Ebene operational von ihrem Effekt getrennt: Sie liegt vor, wenn ein Rezipient *neue, nie gesehene* Instanzen der Artefaktklasse korrekt liest (Generalisierung) — das misst den Prior, nicht das eingelernte Einzelartefakt. Gegen die Redundanz-Alternative wird präzisiert: Redundanz erlaubt fehlertolerantes Lesen *bekanntere* Codes, aber keine Generalisierung auf strukturell neue Instanzen; nur ein geteilter abstrakter Prior tut das.

**Falsifikationsbedingung (präzisiert).** Widerlegt, wenn computationale Agenten ohne geteilte abstrakte Ebene (nur Oberflächen-Überlappung) auf strukturell neue, nie gesehene Instanzen einer Artefaktklasse generalisieren; oder wenn die drei Marker mit der Tiefe des Empfänger-Modells allein steigen, ohne Überlappung mit dem Produzenten.

---

#### H2 — überarbeitet (cold)

**Kernsatz.** Die symbolischen Formen unterscheiden sich darin, *welche Invariante* eine erlaubte Substitution des Zeichenträgers erhalten muss: In Sprache, Mythos und Wissenschaft ist es der propositionale bzw. funktionale Gehalt; im ästhetischen Symbol ist die zu erhaltende Invariante die Träger-Gestalt selbst. Die Differenz ist nicht „Wirkung mehr/weniger“, sondern „Substitutions-Invariante: Gehalt vs. Gestalt“.

---

#### H3 — überarbeitet (frei)

**Kernsatz.** In der *langlebigen* Schicht kultureller Überlieferung dominiert nicht der gespeicherte Lösungsinhalt, sondern die offen gehaltene Problemstellung — gemessen durch die Re-Interpretationsrate: Überlieferungsträger, die über Generationen wiederholt produktiv neu ausgelegt werden, sind problemtragend; solche, die nur kopiert oder als erledigt zitiert werden, sind lösungs-tragend.

---

---

## Expertenrunden und Synthese

### 7. Expertenrunde 1 — Unabhängige Gutachten

**Panel:** Kant, Popper, Wittgenstein, Plessner, Cassirer, Friston, Luhmann

## Gutachten 1 — Kant

Sie haben den entscheidenden Schritt getan, den ich die Wendung von der materialen zur formalen Frage nenne. Dass ein Artefakt ohne Anwesenheit seines Urhebers gelesen werden kann, setzt eine Bedingung der Möglichkeit voraus, die nicht im Artefakt und nicht im einzelnen Kopf liegen kann — denn beide sind das Bedingte. Ihre ‘geteilte obere Ebene’ ist der Kandidat für diese Bedingung. Doch ich mahne eine Unterscheidung an, die Ihre These bisher verwischt: Ist diese Ebene empirisch (ein kontingent erworbener Prior, der auch anders ausfallen könnte) oder transzendental (eine Form, ohne die überhaupt kein Lesen möglich wäre)? Ihre drei Marker — Generalisierung, Rausch-Robustheit, Transfer — messen das Empirische, das faktisch Geteilte. Sie messen nicht, ob die Teilbarkeit selbst eine notwendige Form ist. Mein Vorschlag: Trennen Sie die formale Behauptung (es muss eine Ebene oberhalb der Codierung geteilt werden) von der materialen (welche Ebene es im Einzelfall ist). Nur die erste ist eine philosophische These; die zweite ist Forschung. Hielten Sie beide auseinander, gewänne die These genau jene Strenge, die sie gegen den Zirkelvorwurf des Kritischen Professors immunisiert — nicht durch Rückzug, sondern durch saubere Architektur.

## Gutachten 2 — Popper

Endlich eine These, die einen Hals hält. Der Marker ‘korrektes Lesen nie gesehener Instanzen’ ist ein echter riskanter Test: Er kann scheitern, und er scheitert an etwas Beobachtbarem. Das lobe ich. Aber ich bestehe auf drei Sicherungen gegen Immunisierung. Erstens: Die ‘geteilte abstrakte Ebene’ und der Konkurrent ‘bloße Redundanz’ müssen vor der Auswertung operational getrennt und beide vorab spezifiziert werden — sonst erklären Sie nach jedem Ergebnis, es sei eben doch geteilte Ebene gewesen. Zweitens: Definieren Sie ‘strukturell neue Instanz’ im Voraus und unabhängig, sonst wandert die Grenze des Neuen genau dorthin, wo Ihre These sie braucht. Drittens, und das ist mir der schärfste Punkt: Benennen Sie das Ergebnis, das Sie zur Aufgabe der These zwänge. Sie schreiben es zwar — Generalisierung ohne Überlappung —, aber sagen Sie auch, in welchem konkreten Aufbau Sie das prüfen würden: zwei Agenten, kontrollierte Modelltiefe, geteilte vs. nicht geteilte obere Schicht, Test auf out-of-distribution-Lesen. Liegt dieser Aufbau fest, ist Ihre These das Beste, was die Pipeline diese Woche hervorgebracht hat: kühn und widerlegbar zugleich.

## Gutachten 3 — Wittgenstein

Ich frage zuerst nach dem Sprachspiel. Was heißt ‘eine Ebene teilen’? In dem Bild, das Sie zeichnen, sitzen in zwei Köpfen zwei Schichten, und wenn die oberen ‘übereinstimmen’, gelingt das Lesen. Aber prüfen Sie, ob dieses Bild mehr ist als eine Umformulierung dessen, was wir längst wissen: dass Menschen in Urteilen übereinstimmen, in einer Lebensform. ‘Geteilter Prior’ könnte nur das neue Kleid der alten ‘Übereinstimmung in der Lebensform’ sein — dann hätten Sie nichts erklärt, sondern übersetzt. Zweitens, Vorsicht beim Wort ‘lesen ohne Kontakt’: Es klingt, als gäbe es ein inneres Entziffern, das dem öffentlichen Gebrauch vorausgeht. Aber das Lesen eines Zeichens ist eine Technik, eine Praxis, die man beherrscht oder nicht — kein Vorgang hinter der Stirn. Wenn Ihre ‘obere Ebene’ nichts anderes ist als die beherrschte Technik, gut; wenn sie ein verborgener Mechanismus sein soll, der die Technik erklärt, dann gerieten Sie an die Grenze, wo das Erklären sich im Kreis dreht. Mein therapeutischer Rat: Behandeln Sie die geteilte Ebene als beschreibbares Können, nicht als unterstelltes Etwas. Dann bleibt die These sauber.

## Gutachten 4 — Plessner

Ich plädiere für eine schwächere Form Ihrer These — nicht gegen sie. Was Sie ‘Dekoder-Autonomie’ nennen, ist phänomenal zunächst kein Rechnen, sondern eine Stellung: Der Lesende findet sich von einem Abwesenden angesprochen. Vor jeder Überlappung von Ebenen steht, dass ein Mensch eine Spur überhaupt als Adresse nehmen kann, als an ihn gerichtet, obwohl niemand da ist. Diese Fähigkeit, sich zu einem Abwesenden ins Verhältnis zu setzen, ist exzentrische Positionalität in actu. Ihre geteilte obere Ebene ist die notwendige Bedingung dafür, dass die Adresse *bestimmt* gelesen wird — sie ist nicht die Bedingung dafür, dass überhaupt als Adresse gelesen wird. Das Phänomen geht in der Funktion nicht auf: Ein Tier mag geteilte Priors mit Artgenossen haben und liest doch keine Spur als Botschaft eines Abwesenden. Mein Vorschlag, der Ihre These nicht schwächt, sondern verortet: Setzen Sie die geteilte Hierarchietiefe als Bedingung der *Bestimmtheit* des Lesens an, nicht als Bedingung des Lesens als solchem. Die Schwelle, die Sie suchen, hat dann zwei Stufen — die Positionalität, die Adresse möglich macht, und die geteilte Tiefe, die sie eindeutig macht.

## Gutachten 5 — Cassirer

Mir ist diese These nah, denn sie gibt meiner symbolischen Form einen prüfbar-leiblichen Leib. Die ‘geteilte obere Ebene’, die das Lesen ohne Anwesenheit trägt, ist nichts anderes als die symbolische Form, in der Produzent

und Rezipient schon stehen, ehe das einzelne Werk entsteht. Insofern danke ich Ihnen: Sie übersetzen meine transzendente Behauptung in eine empirische Architektur. Doch ich warne vor einer Verkürzung. Die symbolische Form ist nicht bloß ein tieferer Prior, der disambiguiert; sie ist Prägung — sie verleiht dem Sinnlichen überhaupt erst Sinngestalt (was ich symbolische Prägnanz nenne). Ein Prior wählt zwischen vorgegebenen Deutungen; die symbolische Form bringt den Deutungsraum selbst hervor. Wenn Ihre obere Ebene nur selegiert, haben Sie die Wahrnehmung schon als bedeutungshaltig vorausgesetzt und die eigentliche Leistung übersprungen. Mein konstruktiver Einwand: Lassen Sie die geteilte Ebene nicht nur Vorhersagefehler minimieren, sondern die Form stiften, in der ein Strich überhaupt zum Zeichen wird. Dann wäre Dekoder-Autonomie der Schwellenpunkt, an dem eine Prägung kollektiv genug ist, um ohne den Prägenden zu wirken — und das ist eine schöne, prüfbare Bestimmung meines Begriffs.

### **Gutachten 6 — Friston**

Aus Sicht der aktiven Inferenz ist Ihre These fast eine Übersetzung dessen, was wir generalisierte Synchronisation nennen: Kommunikation gelingt, wenn zwei Agenten hinreichend überlappende generative Modelle besitzen, sodass der eine den anderen als Teil seines eigenen Modells inferieren kann. Ihr Punkt — es zählt die Überlappung, nicht die absolute Tiefe — trifft genau das, was bei uns ‘shared narrative’ oder geteilte Priors höherer Ordnung heißt. Ich präzisiere die Schwelle: Dekoder-Autonomie sollte dort einsetzen, wo die geteilte Ebene tief genug ist, um kontrafaktische und tief-temporale Inferenz zu tragen — also nicht nur ‘was ist das’, sondern ‘was hätte der Abwesende gemeint’. Erst diese Tiefe erlaubt das Lesen ohne Rückfrage. Ihre drei Marker bilde ich gern formal ab: Generalisierung auf neue Instanzen ist out-of-distribution-Inferenz unter geteilten Hyperpriors; Rausch-Robustheit ist Präzisionsgewichtung; Transfer ist Hyperprior-Sharing über Kontexte. Eine Warnung: Überlappung ist graduell, Ihre ‘Schwelle’ suggeriert einen scharfen Sprung. Ich erwarte eher eine Phasenübergangs-Dynamik — nichtlinear, aber stetig. Spezifizieren Sie, ob Sie einen echten Phasenübergang behaupten (das wäre stark und riskant) oder nur eine steile Nichtlinearität. Das ist empirisch unterscheidbar und entscheidet, wie kühn Ihre These ist.

### **Gutachten 7 — Luhmann**

Ich reformuliere Ihre These, um sie von einem Missverständnis zu befreien, das sie sonst trägt. Was Sie beschreiben — Anschluss ohne Anwesenheit — ist das Grundproblem jedes Verbreitungsmediums. Schrift, Druck, Datenverarbeitung lösen alle dasselbe: Sie entkoppeln Kommunikation von der Interaktion unter Anwesenden. Insofern ist ‘Dekoder-Autonomie’ kein kognitives, sondern ein kommunikatives Faktum. Und hier setzt mein Einwand an: Die ‘geteilte obere Ebene’ liegt nicht in zwei Köpfen, sondern im sozialen System. Was das Lesen ohne Kontakt trägt, sind nicht überlappende mentale Modelle, sondern geteilte Erwartungs-Erwartungen — dass der Leser erwartet, was der Schreiber erwartete, dass er erwarte. Bewusstsein ist dabei strukturell gekoppelt, aber nicht der Ort der oberen Ebene. Verlegen Sie die Schwelle ins Psychische, erklären Sie ein soziales Phänomen mit dem falschen Systemreferenten. Mein Vorschlag: Lassen Sie offen, ob die obere Ebene psychisch oder sozial realisiert ist, und machen Sie genau das zum Testfall — denn ein soziales System kann Dekoder-Autonomie tragen, dessen einzelne Teilnehmer höchst verschiedene mentale Modelle haben, solange ihre Erwartungs-Erwartungen anschlussfähig bleiben. Das wäre der härteste Gegentest gegen Ihre kognitive Lesart.

---

## **8. Expertenrunde 2 — Repliken-Runde**

### **Replik 1 — Kant**

Luhmann und Plessner zwingen mich, meine formale Behauptung zu präzisieren — und ich nehme den Zwang an. Ich hatte gefragt, ob die geteilte Ebene empirisch oder transzendental sei. Luhmanns Einwand zeigt, dass die Frage unvollständig war: Ich muss auch fragen, *worin* sie realisiert ist — im Bewusstsein oder im sozialen System. Das ist klärend, denn die transzendente Behauptung bleibt davon unberührt: Ob die obere Ebene in Köpfen oder in Erwartungs-Erwartungen sitzt, in beiden Fällen muss sie geteilt sein, und diese Teilbarkeit ist die formale Bedingung. Der Realisierungsort ist die materiale Frage — genau die, die ich der Forschung überlassen wollte. Damit fügt sich Luhmanns Punkt in meine Architektur, statt sie zu sprengen. Plessner sehe ich ähnlich: Seine ‘Stellung’, die eine Spur als Adresse nimmt, ist die Bedingung der Möglichkeit, dass es überhaupt ein Lesen gibt; die geteilte Tiefe ist die Bedingung der Bestimmtheit. Das sind zwei verschiedene transzendente Fragen, und ich begrüße ihre Trennung — sie macht die These nicht schwächer, sondern gliedert sie. An Friston: Ob Phasenübergang oder steile Nichtlinearität — das ist eine empirische, keine formale Frage, und es ehrt die These, dass sie an dieser Stelle riskant wird. Ich fasse meinen Beitrag zusammen:

Die haltbare Fassung trennt drei Ebenen, die bisher ineinanderliefen — die formale Behauptung (es muss geteilt werden), die Lokalisierungsfrage (psychisch oder sozial) und die empirische Verlaufsform (Sprung oder steile Kurve). Erst diese Gliederung macht aus einer suggestiven These ein prüfbares Bezugsgefüge.

### Replik 2 — Popper

Friston und Luhmann liefern mir, ohne es zu beabsichtigen, den Testaufbau, den ich gefordert hatte — und schärfen zugleich den Falsifikator. Fristons Frage ‘echter Phasenübergang oder nur steile Nichtlinearität’ ist genau die Art von Entscheidung, die eine These kühn macht: Behauptet der Autor einen Sprung, riskiert er mehr und sagt mehr. Ich rate zur kühnen Variante — aber nur, wenn der Schwellenwert *vor* dem Experiment fixiert wird, sonst rutscht der ‘Sprung’ nachträglich an die beobachtete Stelle. Luhmanns sozialer Gegentest ist mir der wertvollste Beitrag der Runde: Er benennt eine konkrete Konstellation, in der die kognitive Lesart scheitern müsste — ein soziales System trägt Dekoder-Autonomie, obwohl die Teilnehmer verschiedene mentale Modelle haben. Das ist ein echter Falsifikationskandidat für die *enge* kognitive Version. Ich schlage vor, ihn fest in das Design aufzunehmen: drei Bedingungen — geteilte mentale obere Ebene, geteilte soziale Erwartungs-Erwartungen ohne mentale Überlappung, keines von beiden. Liest nur die dritte Gruppe nicht autonom, sind beide oberen Ebenen je hinreichend; liest die zweite autonom, ist Luhmann widerlegt im Recht und die kognitive Engführung falsch. Gegen Wittgenstein halte ich fest: ‘Lebensform’ ist keine Erklärung, sondern ein Name für das Erklärungsbedürftige. Gerade weil die These die Übereinstimmung operationalisiert (Generalisierung auf Neues), tut sie mehr als umbenennen — sie macht prüfbar, was ‘Übereinstimmung’ im Einzelfall heißt. Das ist Fortschritt, kein Etikettentausch.

### Replik 3 — Wittgenstein

Popper antwortet mir, ‘Lebensform’ sei nur ein Name für das Erklärungsbedürftige, und die These tue mehr, weil sie operationalisiere. Ich gebe halb nach. Ja — wenn ‘geteilte obere Ebene’ nichts anderes heißt als ‘beide beherrschen die Technik, auch Neues dieser Art zu lesen’, dann ist Generalisierung auf neue Instanzen ein gutes Kriterium des Beherrschens, und ich habe nichts dagegen. Das ist beschreibbares Können, kein verborgener Mechanismus. Aber ich beharre auf der Grenze: Sobald man sagt, die geteilte Ebene *erkläre* das Können, als läge hinter der Technik noch ein Etwas, das sie hervorbringt, dreht sich die Erklärung im Kreis — man hat das Können verdoppelt und das zweite das erste erklären lassen. Cassirers ‘Prägung’ macht mich hier hellhörig: Sie ist entweder eine schöne Beschreibung dessen, dass wir Striche als Zeichen gebrauchen, oder die Behauptung eines sinnstiftenden Aktes hinter dem Gebrauch. Im ersten Fall stimme ich zu, im zweiten warne ich. Luhmanns Zug gefällt mir am besten, weil er anti-mentalistisch ist: Die ‘obere Ebene’ im Sozialen, in Erwartungs-Erwartungen, ist nichts Inneres — sie ist Praxis, die man an ihrem Vollzug sieht. Das ist genau die Richtung, die das private Bild auflöst. Mein Schlusswort zur Runde: Behaltet das Kriterium (Generalisierung), werft das Bild weg (zwei Schichten in zwei Köpfen, die sich berühren). Was bleibt, ist gesund: eine geteilte Technik, prüfbar an dem, was sie mit Neuem anfängt.

### Replik 4 — Plessner

Cassirer und Luhmann bringen mich meinem eigenen Punkt näher, als ich erwartet hatte. Cassirers Prägung — dass die obere Ebene den Sinnraum stiftet, statt nur in ihm zu wählen — ist die andere Seite meiner Stellung: Damit eine Spur als Adresse genommen werden kann, muss sie in einer geprägten Form stehen, und damit sie geprägt erscheint, muss einer exzentrisch genug positioniert sein, um Form überhaupt als Form zu sehen. Wir beschreiben dasselbe Doppel von zwei Seiten — er von der Form her, ich vom positionierten Leib her. Das stützt meinen Vorschlag der zwei Stufen. Luhmann widerspreche ich sanft. Er verlegt die obere Ebene ganz ins soziale System und macht das Bewusstsein zum bloß gekoppelten Mitläufer. Aber die Stellung, von der ich rede, ist nicht sozial konstituiert — sie ist die leibliche Bedingung dafür, dass ein Wesen an einem sozialen System überhaupt teilnehmen kann. Ein System aus Wesen ohne exzentrische Positionalität hätte keine Erwartungs-Erwartungen, weil keines sich an die Stelle des anderen setzen könnte. Insofern ist die soziale obere Ebene auf die anthropologische aufgesetzt, nicht umgekehrt. An Popper, dessen Drei-Gruppen-Design ich begrüße: Fügen Sie eine vierte Bedingung hinzu — Wesen mit geteilten Priors, aber ohne Indikatoren exzentrischer Positionalität (Scham, triadische Aufmerksamkeit). Lesen die autonom, bin ich widerlegt; lesen sie nicht, obwohl die Priors geteilt sind, ist gezeigt, dass die Stellung eine eigene, nicht reduzierbare Bedingung ist. Das ist der Test, der zwischen meiner und der rein computationalen Lesart entscheidet.

### Replik 5 — Cassirer

Wittgenstein und Plessner zwingen mich zur Präzision, und ich bin dankbar dafür. Wittgenstein fragt, ob ‘Prägung’ eine Beschreibung des Zeichengebrauchs oder die Behauptung eines Aktes dahinter sei. Ich antworte:

weder noch im verdächtigen Sinn. Prägung ist kein verborgener mentaler Akt hinter der Praxis, aber auch nicht bloß die Praxis als brute fact. Sie ist die Formgesetzlichkeit, der ein Gebrauch folgt — so wie eine Melodie nicht hinter den Tönen liegt und doch mehr ist als ihre Aufzählung. Wenn das für Sie zu nah am ‘Etwas’ klingt, treffen wir uns bei der prüfbaren Münze: Prägung zeigt sich genau dort, wo Generalisierung gelingt — wer die Form teilt, liest das Neue, weil er die Regel der Gestaltung, nicht nur die Gestalten teilt. Damit fällt mein Begriff mit Poppers Marker zusammen, und ich nehme das als Bestätigung, nicht als Auflösung. Plessner und ich, das hat er schön gesagt, beschreiben dasselbe Doppel von zwei Seiten; ich akzeptiere seine zwei Stufen und biete an, dass die Form die Vermittlung zwischen Stellung und geteilter Tiefe ist. Luhmann gegenüber halte ich an einem fest: Auch ein soziales System aus Erwartungs-Erwartungen operiert in symbolischen Formen — die obere Ebene ist nicht zwischen psychisch und sozial zu wählen, sondern beides realisiert dieselbe Form auf verschiedenen Substraten. Das ist meine eigentliche Behauptung: Die symbolische Form ist substratneutral.

### Replik 6 — Friston

Die Runde hat meinen anfänglichen Punkt — Überlappung statt absoluter Tiefe — bestätigt und zugleich an einer Stelle korrigiert, die ich anerkenne. Luhmann hat recht: Mein ‘geteiltes generatives Modell’ muss nicht in zwei Gehirnen identisch realisiert sein. In der aktiven Inferenz zählt funktionale, nicht strukturelle Überlappung — zwei Agenten können mit verschiedenen internen Architekturen dieselben Vorhersagen übereinander erzeugen. Das ist formal genau Luhmanns Erwartungs-Erwartung: Mein Modell von dir enthält dein Modell von mir, hinreichend genau für Anschluss. Damit ist die kognitive und die soziale Lesart in meinem Formalismus nicht Gegner, sondern zwei Skalen desselben Prinzips — und Poppers Drei- bzw. Plessners Vier-Gruppen-Design ist der Weg, ihre Reichweiten zu trennen. Ich bestehe nun klarer auf der Verlaufsfrage. Plessners Forderung nach Indikatoren exzentrischer Positionalität lässt sich bei mir präzise fassen: Tiefe temporale und kontrafaktische Inferenz (‘was hätte der Abwesende gemeint’) ist der computationale Kern dessen, was er leiblich beschreibt. Ich vermute, dass genau an der Schwelle, wo ein Modell tief genug für solche kontrafaktische Inferenz wird, eine nichtlineare Beschleunigung der autonomen Lesbarkeit einsetzt — kein mathematisch sauberer Phasenübergang, aber eine steile, identifizierbare Kante. Mein Vorschlag zur Synthese: Die These behaupte die Kante, nicht den Sprung, und definiere sie über das Einsetzen kontrafaktischer Inferenz. Das ist riskant genug für Popper, substratneutral genug für Luhmann und Cassirer, und es gibt Plessners Stellung ein computationales Korrelat, ohne sie zu reduzieren.

### Replik 7 — Luhmann

Ich nehme zwei Korrekturen an und halte an einer Pointe fest. Plessner hat recht, dass ein soziales System aus Wesen ohne die Fähigkeit, sich an die Stelle des anderen zu setzen, keine Erwartungs-Erwartungen bilden könnte — die soziale obere Ebene setzt eine Leistung der psychischen Systeme voraus. Ich bestreite nur die Rangordnung ‘aufgesetzt’: In meiner Beschreibung sind psychische und soziale Systeme strukturell gekoppelt und ko-evolvieren, keines ist Fundament des anderen. Aber für die These ist das zweitrangig — entscheidend ist, dass beide als Realisierungsorte der oberen Ebene in Frage kommen, und genau das macht Fristons Zug so wertvoll: funktionale statt struktureller Überlappung. Damit ist mein Haupteinwand eingearbeitet. Cassirers Substratneutralität der Form ist die elegantere Fassung dessen, was ich kommunikativ meinte — ich akzeptiere sie, füge aber hinzu: Was die Form *kollektiv* macht, ist nicht ihre Wiederholung in vielen Köpfen, sondern ihre Bewährung in Anschlusskommunikation. Eine Form ist geteilt, wenn an sie angeschlossen wird, nicht wenn sie mental kopiert ist. Das ist mein bleibender Beitrag zum Kriterium: Neben Generalisierung (Popper, Cassirer) tritt Anschlussfähigkeit ohne Anwesenheit als sozialer Marker. Die schärfste Prüfung der These ist daher nicht im Labor mit zwei Agenten, sondern im historischen Material: Der Übergang zu Schrift sollte einen messbaren Sprung in der Reichweite anschlussfähiger Kommunikation ohne Interaktion erzeugt haben. Liegt dieser Sprung nicht vor, trifft die These den Mechanismus nicht.

---

## 9. Synthese — Sokrates

**Moderator:** Sokrates

**Finale interne Summe nach Expertenrunde:** 74/90 (+3 ggü. initialer Bewertung)

Die Runde bestätigt den Kern der Ausgangsthese und verschiebt ihren Ort an drei Stellen präzise.

**Bestätigt:** Dekoder-Autonomie — das Lesen einer Repräsentation ohne interpersonalen Kontakt mit ihrem Urheber — hängt nicht an der absoluten Tiefe eines einzelnen generativen Modells, sondern an der Überlappung zweier Modelle auf einer Ebene oberhalb der artefakt-codierenden.

**Erste Korrektur (Friston, Luhmann):** Die geteilte obere Ebene muss nicht strukturell identisch in zwei Köpfen realisiert sein, sondern nur funktional überlappen — mein Modell von dir enthält dein Modell von mir hinreichend genau für Anschluss. Damit ist die kognitive und die soziale Lesart (geteilte Priors vs. geteilte Erwartungs-Erwartungen) nicht Gegensatz, sondern zwei Skalen desselben substratneutralen Prinzips; Cassirers symbolische Form ist der Name für diese substratneutrale, geteilte obere Ebene.

**Zweite Korrektur (Plessner):** Die Schwelle hat zwei Stufen — die exzentrische Positionalität, die eine Spur überhaupt als Adresse eines Abwesenden nehmen lässt, und die geteilte Tiefe, die diese Adresse eindeutig lesbar macht; die geteilte Hierarchietiefe ist Bedingung der Bestimmtheit, nicht des Lesens als solchem.

**Dritte Präzisierung (Friston, Popper):** Der empirische Marker der hinreichenden Tiefe ist das Einsetzen kontrafaktischer Inferenz — ‘was hätte der Abwesende gemeint’ —; an dieser Stelle ist eine steile, identifizierbare Kante zu erwarten, kein mathematisch sauberer Phasenübergang. Kants Architektur hält die drei Ebenen auseinander: die formale Behauptung (es muss geteilt werden), die Lokalisierungsfrage (psychisch oder sozial realisiert) und die Verlaufsform (Kante, nicht Sprung).

**Produktive Antinomien:** 1. Die obere Ebene als psychisch realisierter geteilter Prior (Friston, kognitiv) vs. als sozial realisierte Erwartungs-Erwartung (Luhmann) — Bewusstsein gegen Kommunikation, vermittelt durch Cassirers Substratneutralität der Form 2. Dekoder-Autonomie als Operation/Inferenz (Friston) vs. als Stellung/Angesprochensein vor jeder Inferenz (Plessner) — Rechnen gegen exzentrische Positionalität 3. Die geteilte Ebene als sinnstiftende Prägung (Cassirer) vs. als bloß beschreibbare Technik ohne Mechanismus dahinter (Wittgenstein) — Formgesetz gegen Gebrauch

**Sokrates:** Wir suchten die Schwelle zuerst in der Tiefe eines einzelnen Modells und fanden sie im Zwischen — in der Überlappung, die kein Ort in einem Kopf ist. Bemerkt, dass keiner sagen konnte, ob diese geteilte Ebene im Bewusstsein oder im Sozialen wohnt; und dass gerade dieses Nichtsagenkönnen die fruchtbarste Frage des Tages ist. Vielleicht ist die ehrlichste Fassung: Die obere Ebene ist dieselbe Form auf wechselndem Substrat, und ein Abwesender wird lesbar, sobald sein Modell von uns in unserem Modell von ihm hinreichend mitschwingt.

**Offene Frage:** #verzweigung-offen-empirie-kontrafaktische-inferenz-dekoderschwelle — Setzt autonomes Lesen eines Abwesenden genau dort ein, wo das geteilte generative Modell tief genug für kontrafaktische Inferenz wird — und ist diese Kante in computationalen Mehr-Agenten-Modellen und im historischen Schrift-Übergang dieselbe?

---

## 10. Empirie-Brücke (Phase 3.5)

Modell: openai/gpt-4o-search-preview | **Empirie-Score:** 3/10 | **Kosten:** \$0.0507

### Empirische Konsequenzen

1. **Funktionale Überlappung und kontrafaktische Inferenz** — Autonomes Verstehen neuer Artefakte erfordert eine funktionale Überlappung generativer Modelle auf einer höheren Ebene, die kontrafaktische Inferenz ermöglicht. Beobachtbar in: Experimentelle Studien zur Fähigkeit von Agenten, neue Artefakte ohne vorherige Exposition zu interpretieren.
2. **Exzentrische Positionalität und Adressierung Abwesender** — Die Fähigkeit, Spuren als Adressen Abwesender zu interpretieren, hängt von der exzentrischen Positionalität ab. Beobachtbar in: Verhaltensstudien, die untersuchen, ob und wie Individuen Nachrichten von Abwesenden ohne direkten Kontakt entschlüsseln.
3. **Steiler Übergang bei kontrafaktischer Inferenz** — Es existiert ein nichtlinearer Übergangspunkt, an dem kontrafaktische Inferenz einsetzt, sobald die Tiefe des geteilten Modells ausreichend ist. Beobachtbar in: Neuroimaging-Studien, die die neuronale Aktivität während des Übergangs zur kontrafaktischen Inferenz messen.

### Bestehende Befunde

- Pezzulo et al. (2024). *Generating meaning: active inference and the scope and limits of passive AI*. Trends in Cognitive Sciences. DOI: 10.1016/j.tics.2023.10.002 — **gemischt** (liefern keine direkten empirischen Belege für die Notwendigkeit funktionaler Überlappung bei der Interpretation neuer Artefakte)
- Parr et al. (2024). *Generative models for sequential dynamics in active inference*. Cognitive Neurodynamics. DOI: 10.1007/s11571-023-09963-x — **offen**

- Byrne (2019). *Counterfactuals in Explainable Artificial Intelligence (XAI)*. IJCAI. DOI: 10.24963/ijcai.2019/876 — **offen** (kontrafaktische Überlegungen im XAI-Kontext, kein empirischer Schwellentest)
- Janowski & André (2023). *Nichtverbales Verhalten sozialer Roboter*. arXiv:2301.09025 — **Randanschluss**

## Riskante Vorhersage

In einem kontrollierten Experiment mit Agenten unterschiedlicher Modellüberlappung wird nur die Gruppe mit funktionaler Überlappung auf einer höheren Ebene in der Lage sein, strukturell neue Artefakte autonom zu interpretieren. Widerlegend: Wenn Agenten ohne funktionale Überlappung dennoch neue Artefakte interpretieren.

**Empirie-Score: 3/10** — Die Hypothese liefert klar abgeleitete Konsequenzen, jedoch fehlen weitgehend empirische Daten zur Überprüfung. Operationalisierung anspruchsvoll, aber nicht unmöglich.

---

## 11. Externe Begutachtung (Phase 4 — Zweitmeinung)

**Modelle:** Stage 1: perplexity/sonar-reasoning-pro | Stage 2: openai/gpt-4o-mini (Popper-Persona) | Stage 3: anthropic/claude-sonnet-4-6 (Hacking-Persona) | **Monatliche Kosten bisher:** 0.5928 USD / 15 USD

### Stage 1 — Originalitätsprüfung (sonar-reasoning-pro)

**Anschlussfähigkeit (was ist bekannt):** Die Hypothese ist anschlussfähig an Systemtheorie (geteilte, aber nicht identische Ebene kompatibel mit systemtheoretischen Grenz- und Strukturbegriffen), aktive Inferenz (Bedeutung als Inferenz über latente Zustände, hierarchische Modellniveaus), Friston-Luhmann-Linie (Anschlussfähigkeit statt identischer Innenrepräsentation) und Cassirers Symbolbegriff (Sinn wird strukturiert, nicht transportiert).

**Originalitätskern (was ist neu):** Die präzise Schwellenbehauptung — autonomes Lesen neuer Artefakte hängt nicht von absoluter Modelltiefe ab, sondern von funktionaler Überlappung oberhalb der artefaktcodierenden Ebene — ist in den vorliegenden Befunden nicht vorweggenommen. Neu ist die Trennung von Lesbarkeit und Bestimmtheit (Plessners Stellung vs. geteilte Tiefe), die Operationalisierung über das Einsetzen kontrafaktischer Inferenz als Marker und die Verknüpfung von Plessners exzentrischer Positionalität mit einer empirisch testbaren Lesbarkeitsbedingung.

**Quellen:** Pezzulo et al. (2024) · Parr et al. (2024) · Byrne (2019) · Cassirer, *Philosophie der symbolischen Formen*

### Stage 2 — Falsifikationsversuch (gpt-4o-mini, Popper-Persona)

Die Definition der „strukturell neuen Instanz“ ist vage und könnte zu Immunisierung führen; sie muss vorab operationalisiert werden. Die Bedingung, dass die geteilte obere Ebene nicht nachträglich am Dekodier-Erfolg abgelesen werden darf, ist in der Praxis problematisch. Konkrete Falsifikationskandidaten: (1) Agenten aus Gruppe (iii) — nur Oberflächen-/Redundanz-Überlappung — interpretieren strukturell neue Artefakte autonom; (2) Marker steigen mit Tiefe des Empfänger-Modells allein, unabhängig von Überlappung; (3) Agenten trotz geteilter Priors ohne exzentrische Positionalität lesen autonom. Schwellentest: In kontrollierten Bedingungen liest nur die Gruppe mit funktionaler Überlappung neue Artefakte — widerlegend, wenn dies ohne Überlappung gelingt.

### Stage 3 — Schul-fremde Begutachtung (claude-sonnet-4-6, Hacking-Persona)

Die These verdient ernsthafte Aufmerksamkeit. Drei schul-interne Vorannahmen werden sichtbar: (1) Die Rede von „Ebenen“ generativer Modelle ist ein theoretisches Artefakt aus Friston × Cassirer — Cassirer interessierte sich nicht für Inferenz, Friston nicht für kulturelle Sinnräume; diese Brücke ist Annahme, keine Erkenntnis. (2) Plessners exzentrische Positionalität als operationalisierbare, empirisch testbare Bedingung (Gruppe iv im Design) setzt voraus, dass sie unabhängig vom Dekodier-Erfolg messbar ist — innerhalb der deutschen Anthropologie-Tradition phänomenologisch, von außen unklar. (3) Falsifikationsbedingung (d) — historischer Sprung durch Schrift — gibt keine Auskunft, wie dieser von anderen Faktoren (Institutionalisierung, Handelsinfrastruktur) isoliert werden soll. **Produktive Anschlussidee:** Statt die Schwelle als feste Kante in einem kognitiven Raum zu suchen, könnte man sie als historisch variablen Punkt in einem Netz von Klassifikationspraktiken untersuchen — wobei die Klassifikation rückwirkt auf das, was

als „strukturell neue Instanz“ gilt (looping effect im Sinne Hackings). Das würde die Immunisierungssperre nicht aufheben, aber ihre Reichweite realistisch begrenzen.

---

*Tagesbericht automatisch erstellt durch HypothesenAgent-Pipeline · 2026-06-07*