

Inhaltsverzeichnis

Hypothesentag-Gutachten	1
Die Gewinnerthese	1
Initiale Fassung	1
2. Verzweigungs-Pickup	1
Reformulierte Fassung (nach Kritischem Professor)	2
Expertenrunden und Synthese	2
7. Bewertung mit dreifacher Klassifikation	2
Dreifache Klassifikation	2
H1 — Der nicht-eliminierbare Hyperprior als Geltungsort (empirisch_pruefbar)	3
H2 — Das Demarkationsproblem des Vertrauens (begriffsanalytisch)	3
H3 — Sphärenspezifische Normativität als Koordinations-Problem (methodologisch_wissenschaftstheoretisch)	4
Auswahl	4
Reservoir-Profile (nicht gewählt)	5
8. Expertenrunde 1 — unabhängige Gutachten	5
9. Expertenrunde 2 — Repliken-Runde	7
10. Synthese (Sokrates)	9
Synthese im Sokrates-Modus	9
Finale Hypothese	10
Finale Bewertung mit Begründung	10
Lerneffekt der Pipeline	11
Frage an die nächste Runde	11
11. Reservoir-Verweise	11
11.5 Empirie-Brücke (Phase 3.5)	12
Empirie-Brücke (Phase 3.5, Claude mit Websuche)	12
Empirische Konsequenzen	12
Bestehende Befunde	12
Riskante Vorhersage (Schwellentest)	13
Offene empirische Fragen	14
Empirie-Score	14
12. Externe Begutachtung (Phase 4)	14

Hypothesentag-Gutachten

Die Gewinnerthese

In jedem endlichen hierarchischen FEP-Modell einer normativen Praxis ist der Top-Level-Hyperparameter, der festlegt, was als gelungene Form zählt, exogen zur generativen Dynamik — gesetzt, nicht aus den Systeminvarianten gewonnen. Diese Exogenität ist die notwendige (nicht hinreichende) empirische Signatur der Grenze Genese/Geltung (Brücke B): der Ort, an dem das Modell ein Erfolgskriterium voraussetzt, das es selbst nicht erzeugt.

Initiale Fassung

2. Verzweigungs-Pickup

- **warm_pick** — empirie-parameterwahl-in-fep-modellen [Makro-Thema Erkenntnistheorie_Cassirer], offene Verzweigung aus [[06 Hypothesentag/2026-06-16]]. Trägt H1.
- **cold_pick** (Diversitäts-Pflicht) — Luhmann - Vertrauen als Komplexitätsreduktion [[04 Ressourcen/Philosophen & Werke/Luhmann - Vertrauen als Komplexitätsreduktion]], virgin node (Makro-Thema Systemtheorie_Luhmann, noch nie verarbeitet). Trägt H2. Generation-Directive: im Luhmann-Vokabular argumentieren, nicht in Sollwert/Regelkreis/Active-Inference umdichten.
- **devil_advocate_pick** — nicht aktiviert (Dienstag, kein Echo-Chamber-Alarm, keine Drift).
- **reservoir_pick** (Kadenz gefeuert) — Reservoir - Sphärenspezifische Normativität Brückenannahmen 2026-06-02 (Score 69 → effektiv 73, Domäne Soziologie, 28 Tage unbeachtet). Ersetzt H3 (freie Generalisierung/Exploration).
- **zieldomaene H2** — Epistemologie (Kant/Popper/Wittgenstein/Hacking/Quine).

- **exploration_domaine** — Evolutionäre Biologie & Paläontologie (nicht verwendet, da reservoir_pick H3 belegt).

Reformulierte Fassung (nach Kritischem Professor)

Hypothese 1 (warm_pick: empirie-parameterwahl-in-fep-modellen) — Der nicht-eliminierbare Hyperprior als Geltungsort

Kernsatz. In FEP-Modellen kultureller oder normativer Praxis ist der frei gewählte Präzisions- bzw. Komplexitätsparameter, der die „Praxis → Form“-Schwelle markiert, kein kalibrierbarer Restfreiheitsgrad, sondern der systematische Ort, an dem das Erfolgskriterium von außen eintritt: Jede hierarchische Präzisions-Lernschleife (empirical Bayes) endet in einem obersten Hyperprior, der gesetzt und nicht aus Daten gelernt wird — und genau dieser nicht-eliminierbare Hyperprior trägt die Geltungsdimension, die das Modell formal nicht erzeugt.

Begründung. Die offene Verzweigung #verzweigung-offen-empirie-parameterwahl-in-fep-modellen (aus [[06 Hypothesentag/2026-06-16]]) hatte den empirischen Befund festgehalten, dass die Praxis-Form-Schwelle in konkreten FEP-Modellen enaktiver Nischenkonstruktion (Constant et al. 2019, Ramstead et al. 2016) durch einen frei gewählten Parameter markiert ist. Der Standardeinwand (Friston, 06-16) lautete: Präzision sei hierarchisch lernbar, also nicht „frei“. Diese These schärft den Befund gegen genau diesen Einwand: Auch eine hierarchische Lernschleife terminiert in einem endlichen Modell an einem obersten Prior, der nicht gelernt, sondern stipuliert wird; dort — und nur dort — entscheidet sich, was als „gelungene Form“ zählt. Das ruht auf [[Cassirer - Regelung als transzendentes Prinzip]] und dem Gutachten [[Abschlussbericht Cassirer-Friston Fortsetzung]], die die Nicht-Ableitbarkeit der Erfolgsbestimmung als das eigentliche Forschungsdesiderat (Brücke B) benannt haben.

Falsifikationsbedingung. Falsifiziert, wenn ein FEP-Modell einer normativen oder kulturellen Praxis vorgelegt wird, dessen schwellenmarkierende Präzisions-/Komplexitätsparameter sämtlich aus Systeminvarianten abgeleitet sind (vollständig selbstorganisierende oberste Ebene ohne stipulierten Hyperprior) und das die Praxis-Form-Schwelle dennoch reproduziert.

Quelle. warm_pick — verfolgt den Strang aus [[06 Hypothesentag/2026-06-16]] (FEP-Parameterwahl).

Expertenrunden und Synthese

7. Bewertung mit dreifacher Klassifikation

Gewichtung aktiv (Hebel 1): Originalität ×1.6, Tiefe ×1.2, Antinomie-Test ×1.2, Rest ×1.0. Normierungsfaktor 0.882353 auf 90er-Skala. Anker je Methoden-Typ adjustiert (kriterien.yaml).

Dreifache Klassifikation

	Themenfeld	Methoden-Typ	Reichweiten-Klasse
H1	epistemologisch_systemtheoretisch	empirisch_pruefbar (mit methodologisch-wissenschaftstheoretischer Beimischung)	these
H2	erkenntnistheoretisch / soziologische Systemtheorie	begriffsanalytisch (mit epistemologischem Kern)	sondierung
H3	methodologisch / Kulturphilosophie	methodologisch_wissenschaftstheoretisch	sozialwissenschaftlich

H1 — Der nicht-eliminierbare Hyperprior als Geltungsort (empirisch_pruefbar)

Kriterium	Score	Begründung
Originalität	7	Verbindet FEP-Parameterwahl mit dem Geltungsort; über 06-16 hinaus durch operationalisierten Marker (stipuliert vs. geschätzt).
Falsifizierbarkeit	9	Ein Gegenmodell falsifiziert; „Ableitung aus erfolgskriteriumfreiem Invarianzprinzip“ operationalisiert.
Begriffliche Klarheit	8	„Ableitbar“ und „Erfolgsbestimmung“ geschärft; Restdichte im Begriff Erfolgsbestimmung.
Tiefe	8	Berührt Geltung/Genese über eine konkrete Modellierungsbehauptung (max 9 bei diesem Typ).
Forschungsrelevanz	9	Direkter Anschluss an die aktive Active-Inference-Debatte (Friston, Ramstead, Constant).
Interdisziplinäre Anschlussfähigkeit	8	FEP, Philosophie des Geistes, bayes. ML, Cassirer-Epistemologie, Kulturtheorie.
Vault-Anschluss	9	Direkte Vertiefung dichter Knoten (Cassirer-Friston, Brücke B).
Antinomie-Test	7	Friston-Gegenthese (Präzision lernbar) produktiv; Typ-Cap 7.
Publikationsmöglichkeit	7	Philosophie-der-Kognitionswissenschaft-Venue plausibel.
Summe	70/90	Rohsumme 79.2×0.882353 .

H2 — Das Demarkationsproblem des Vertrauens (begriffsanalytisch)

Kriterium	Score	Begründung
Originalität	8	Looping-Effekt (Hacking) auf Luhmanns Vertrauenstheorie — frische Synthese, kein Vault-Vorläufer.
Falsifizierbarkeit	6	Typ-Cap 7; Falsifikation (systeminternes Maß) tragfähig, aber abstrakt.
Begriffliche Klarheit	8	Domäneneinschränkung + epistemisch/konstitutiv getrennt.
Tiefe	8	Berührt die Demarkationsfrage als erkenntnistheoretische Grundfrage.
Forschungsrelevanz	8	Sozialepistemologie des Vertrauens ist aktives Feld.

Kriterium	Score	Begründung
Interdisziplinäre Anschlussfähigkeit	7	Soziologie, Epistemologie, Systemtheorie, polit. Theorie (Typ-Cap 8).
Vault-Anschluss	6	Virgin node, frisch, aber dünn (Einzelknoten).
Antinomie-Test	8	Internalismus vs. Externalismus produktiv (Typ-Cap 10).
Publikationsmöglichkeit	7	Episteme / Synthese.
Summe	65/90	Rohsumme 74.0 × 0.882353.

H3 — Sphärenspezifische Normativität als Koordinations-Problem (methodologisch_wissenschaftlich)

Kriterium	Score	Begründung
Originalität	6	„Dritte Kategorie Koordination“ nah an Reichenbachs Zuordnungsdefinition; Abgrenzung vorhanden, aber Reetikett-Risiko bleibt.
Falsifizierbarkeit	7	Riskante Existenzbehauptung + vorab festgelegte Brücke für (b) (Typ-Cap 8).
Begriffliche Klarheit	8	Koordination / transzendental / empirisch sauber getrennt.
Tiefe	9	Berührt die Schwelle transzendental/empirisch — Grundunterscheidung.
Forschungsrelevanz	7	Cassirer + Moralpsychologie — eher nischig.
Interdisziplinäre Anschlussfähigkeit	7	Wissenschaftstheorie, Cassirer-Forschung, Moral-/Kulturpsychologie.
Vault-Anschluss	8	Reservoir-Reaktivierung; bindet Cassirer-Knoten + Monatsstrang.
Antinomie-Test	8	transzendental vs. empirisch produktiv (Typ-Cap 10).
Publikationsmöglichkeit	7	SHPS / Synthese.
Summe	65/90	Rohsumme 74.0 × 0.882353.

Auswahl

Gewinner: H1 — Der nicht-eliminierbare Hyperprior als Geltungsort (70/90). Höchste Summe, klar vor H2/H3 (65).

Gleichstand H2 = H3 (65): Tiebreak Originalität → H2 (8) > H3 (6). Reihenfolge: H2 zweiter, H3 dritter.

Reichweiten-Klasse Gewinner: these. Eigenständig falsifizierbare Einzelthese mit operationalisiertem Schwellentest (Struktur- + Empiriethese); vertieft den FEP-Parameterstrang in eine riskante Behauptung, verbindet Vault-Knoten (Cassirer-Friston, Brücke B) und öffnet *eine* Folge-Verzweigung (was als „Ableitung aus Invarianten“ zählt) — aber keine mehrlinige Architektur. Daher These, kein Forschungsprogramm.

Panel (epistemologisch_systemtheoretisch): Kant, Popper, Wittgenstein, Plessner, Cassirer, Friston, Luhmann.

Reservoir-Profile (nicht gewählt)

H2 — Demarkationsproblem des Vertrauens (sondierung, 65). Neuer Strang Sozialepistemologie/Looping; Vault-Anker dünn (virgin node), hohe Originalität. Pickup-Kandidat für Tiefe-Prüfung. `forschungsprogramm_kandidat: false`.

H3 — Sphärenspezifische Normativität als Koordinations-Problem (sondierung, 65). Reservoir-Reaktivierung; durch Koordinations-/Duhem-Schärfung + vorab festgelegte Brücke (b) gegenüber Stand 2026-06-02 verbessert. Bleibt produktiv; Original-Reservoir aktualisieren statt neu anlegen.

8. Expertenrunde 1 — unabhängige Gutachten

Panel-Konfiguration. Klassifikation erkenntnistheoretisch / epistemologisch_systemtheoretisch (FEP-Parameterwahl, Geltungsort). Gewählt: `epistemologisch_systemtheoretisch_hypothese`. Begründung: Die These berührt primär die Erkenntnistheorie der Modellierung (Geltung/Genese im FEP), mit systemtheoretischer Beimischung (Luhmann als Beobachter-Korrektiv).

Panel-Mitglieder. Kant, Popper, Wittgenstein, Plessner, Cassirer, Friston, Luhmann.

Gutachten 1 — Kant

Die These setzt fort, was ich seit langem als Heteronomie-Vorbehalt führe: Ein Prinzip, das eine Norm voraussetzt, die es nicht erzeugt, hat seinen Geltungsgrund außerhalb seiner Mechanik. Der nicht-eliminierbare Hyperprior ist genau die Stelle, an der die Gesetzgebung wechselt — von der Regel, die ein System befolgt, zu der Regel, unter der etwas als „gelungen“ überhaupt erst gilt. Soweit Zustimmung. Mein Vorbehalt richtet sich gegen die Vermengung der beiden Kernsätze. (a) ist eine Aussage über die Bedingung der Möglichkeit, dass ein FEP-Modell ein Erfolgskriterium *haben* kann; (b) ist eine empirische Aussage über vorliegende Modelle. Die empirische Beobachtung, dass Autoren ihren Parameter stipulieren, kann (a) niemals begründen — sie kann sie nur illustrieren. Würde man (a) aus (b) ableiten, beginge man genau den Fehler, den die These anderen vorwirft: man bezöge die Geltung der Strukturthese aus der Genese der Modellpraxis. Mein Schärfungsvorschlag: Formuliert (a) ausdrücklich als transzendente Bedingung — „damit ein Modell zwischen gelungener und misslungener Form unterscheiden kann, muss es ein Erfolgskriterium voraussetzen, das es nicht aus seinen Invarianten gewinnt“ — und lasst (b) nur als Bewährung gelten. Der selbstbezügliche Test ist Pflicht: Die These darf ihre eigene Geltung nicht aus dem Erfolg der Pipeline beziehen.

Gutachten 2 — Popper

Den operationalen Marker — der stipulierte statt geschätzte Parameter, dessen Variation die Klassifikation kippt — begrüße ich; endlich etwas, das man an einem konkreten Modell prüfen kann. Doch ich sehe die alte Gefahr wiederkehren: „strukturell nicht ableitbar“ droht zur immunisierten Behauptung zu werden. Jede vorgelegte Ableitung wird der Verteidiger als „schmuggelt das Erfolgskriterium ein“ zurückweisen — und schon ist die These gegen jede Widerlegung gefeit. Deshalb mein harter Einwand: Die These muss *vorab* festlegen, was als „erfolgskriteriumfreies Invarianzprinzip“ zählt. Solange dieser Begriff offen bleibt, beißt die Falsifikationsbedingung nicht. Ich schlage einen konkreten Schwellentest vor: An Constant et al. 2019 und Ramstead et al. 2016 prüfen, ob der präzisionsmarkierende Parameter einer Sensitivitätsanalyse unterzogen wird — wird er es, ist er als frei gewählt ausgewiesen, und (b) ist bestätigt; wird er aus einem deklarierten Symmetrieargument fixiert, ist (b) für dieses Modell widerlegt. Zweitens warne ich vor der bequemen Ruhelage: Eine Formel, der Kant, Cassirer und Friston gleichermaßen zustimmen — „das Erfolgskriterium ist exogen“ —, hat geringen Informationsgehalt, wenn niemand angibt, welcher Befund sie widerlegen würde. Riskant wird die These erst, wenn sie *einen* Modelltyp benennt, in dem ihr Scheitern aussähe. Ohne diese Vorab-Festlegung bleibt sie wahr, aber leer.

Gutachten 3 — Wittgenstein

„Der Parameter trägt die Geltungsdimension“ — hier schiebt sich ein Bild vor die Beschreibung: ein Parameter, der etwas *trägt*, wie ein Pfeiler eine Last. Ich rate, das Bild fallen zu lassen und auf die Verwendung zu sehen. Was die These tatsächlich Klares anbietet, ist die Unterscheidung stipuliert/geschätzt — und das ist kein Bild, sondern eine grammatische Bemerkung, die man am Modell ablesen kann: Der Modellierer *setzt* einen Wert und nennt das Resultat „erfolgreich“. Genau dort liegt der Witz. Doch dann ist Vorsicht geboten: Ist die Nicht-Ableitbarkeit eine Entdeckung über das FEP, oder ist sie eine Bemerkung über unseren Gebrauch des Wortes „Erfolg“? Mein Verdacht: das Zweite. „Aus der Dynamik folgt kein Erfolgskriterium“ sagt dasselbe wie „es gehört nicht zur Grammatik von ‚Dynamik‘, dass sie ein ‚Gelingen‘ enthält“. Die These muss sich entscheiden,

in welchem Sprachspiel sie spielt. Spielt sie das empirische (b), dann ist sie eine Hypothese über Modelle und an Modellen zu prüfen. Spielt sie das grammatische (a), dann gehört sie in die Begriffsklärung und nicht ins FEP — und besteht den Test nur als Klärung, nicht als wahr/falsch-Satz. Mein Schärftungsbeitrag: Beschreibt drei konkrete Modellfälle, in denen ein Modellierer einen Parameter setzt und das Ergebnis „gelungene Form“ nennt. An den drei Fällen zeigt sich, ob ein gemeinsames Muster vorliegt oder bloß eine Redeweise.

Gutachten 4 — Plessner

Ich plädiere, wie so oft, für die schwächere und haltbarere Form. Die FEP-Formalisierung ist notwendige, nicht hinreichende Bedingung — das gehört in den Kernsatz, nicht in eine Fußnote. Der stipulierte Hyperprior ist keine Mauer, die zwei Reiche trennt, sondern die obere Stufe einer Stufung: Unter ihr liegt die lückenlose Dynamik, an ihr beginnt das Urteilen. Mein Einwand gegen die Härte der These: Indem sie alles auf einen Parameter zieht, übersieht sie, dass der Modellierer in einer *Stellung* steht — er erfährt sich als gebunden an ein Kriterium, das er setzt und doch nicht beliebig setzt. Diese erfahrene Verbindlichkeit geht im Parameterwert nicht auf. Sie ist kein Restfreiheitsgrad, sondern das Phänomen, das die These gerade nicht in die Stipulation auflösen sollte. Würde man sie auflösen, verlöre man, was an der Sache anthropologisch interessant ist: dass ein Erfolgskriterium gesetzt wird *und* den Setzenden bindet. Mein Schärftungsvorschlag: Nehmt „notwendige, nicht hinreichende Bedingung“ wörtlich in (a) auf und markiert den Hyperprior als obere Stufe, nicht als Grenzfall. Dann bleibt die These prüfbar und schützt zugleich das Phänomen vor der Reduktion auf eine Zahl im Modell.

Gutachten 5 — Cassirer

Die Nicht-Ableitbarkeit der Erfolgsbestimmung ist, in der Sprache der Parameterschätzung gesagt, der Nachweis, dass die Bedeutungsfunktion nicht aus der Ausdrucksfunktion folgt. Der stipulierte Hyperprior ist die Stelle, an der Darstellung zu Bedeutung wird — wo ein bloßer Wert zum Maßstab des Gelingens erhoben wird. Das ist kein Defekt des Modells, sondern der Ort der symbolischen Stiftung, sichtbar gemacht in der Variationsrechnung. Insofern Zustimmung, und mehr: Die These leistet, was das Cassirer-Friston-Gutachten als Desiderat benannte — sie lokalisiert Brücke B empirisch. Mein Einwand betrifft die Reichweite von (b). Dass *gegenwärtige* Modelle den Parameter stipulieren, zeigt eine Praxis; der eigentliche Punkt ist, dass sie ihn stipulieren *müssen*, weil keine Ausdrucksdynamik eine Bedeutung erzeugt. Die These sollte die Modallücke zwischen „tun“ und „müssen“ nicht der Empirie überlassen. Mein Schärftungsvorschlag: Markiert Brücke B dreifach — formal (Nicht-Ableitbarkeit), symbolisch (Stipulation als Bedeutungssetzung), phänomenal (das erfahrene Maßgeben). Erst diese dreifache Markierung trennt die These von der bloßen Beobachtung, dass Modellierer faul oder pragmatisch sind. Der Hyperprior ist nicht eine Bequemlichkeit, sondern der Index dafür, dass jede Form eine Energie der Bedeutung voraussetzt, die in keiner Substanz, auch nicht in keiner Dynamik, schon enthalten ist.

Gutachten 6 — Friston

I accept the role assigned to my earlier concession but I resist its inflation. Precision is learnable — hierarchically, through empirical Bayes — and much of what looks „freely chosen“ is in fact estimated from data at a lower level. That said: a finite generative model terminates in a top-level prior that is exogenous, and model evidence rewards a parsimony that the dynamics do not themselves derive. So far the thesis is right. My sharp objection concerns the operationalisation. The author rules out a Jeffreys or maximum-entropy prior as „not a derivation of the success criterion“ — but an objective prior *is* derived from an invariance argument, and whether it „presupposes the success criterion“ is precisely the contested point, not a settled one. The whole thesis turns on the phrase „success-criterion-free invariance principle,“ and that phrase is currently doing unearned work. Make it precise or the claim is unfalsifiable. My constructive proposal: state the thesis narrowly as a claim about the *exogeneity of the top-level hyperparameter* in finite hierarchical models of normative practice, and test it on Constant et al. 2019 and Ramstead et al. 2016 — examine whether the top-level precision is fixed by hand or inferred. Do not stake the claim on a structural impossibility of formalising the transition; stake it on the non-inferability of the top-level prior. That is defensible; the stronger claim is not.

Gutachten 7 — Luhmann

Die These sucht, wie alle anderen am Tisch, einen Ort für die Geltung — diesmal einen Parameter, einen Hyperprior. Jeder Ort ist eine Unterscheidung des Beobachters, nicht des Phänomens. Wer den Hyperprior „setzt“, ist der Modellierer, und dass er ihn setzt, ist seine Operation, nicht eine Eigenschaft des Modells. Systemtheoretisch gewendet ist Fristons Konzession präziser als seine Verteidigung: Das FEP kann die

Unterscheidung, mit der es Erfolg zuschreibt — gelungen/misslungen —, nicht zugleich beobachten, während es mit ihr operiert. Das ist sein blinder Fleck. Der stipulierte Parameter ist nicht der „Geltungsort“, sondern die Stelle, an der die Modellierung reflexiv wird: an der die Unterscheidung gelungen/misslungen wieder in das Modell eintritt, das sie verwendet — ein Re-entry. Mein Einwand gegen Kant und Cassirer: Sie geben dem Hyperprior sofort eine Würde — Gesetzgebung, Bedeutungsstiftung. Damit verdoppeln sie nur die Beobachtung erster Ordnung. Mein Schärfungsbeitrag: Fasst den nicht-eliminierbaren Hyperprior als Re-entry der Leitunterscheidung, nicht als anthropologisches Ereignis. Dann wird prüfbar, was sonst feierlich bleibt: Ein Modell, das seinen eigenen Erfolgsmaßstab als Variable führt und zugleich anwendet, erzeugt genau die Paradoxie, die die These als „Nicht-Ableitbarkeit“ beschreibt. Und die These selbst sollte ihre Beobachterposition zweiter Ordnung ausweisen — sonst tut sie, was sie dem FEP vorwirft.

9. Expertenrunde 2 — Repliken-Runde

Verfahren. Jeder Experte hat die sechs anderen Gutachten aus Runde 1 erhalten. Die folgenden Repliken sind Antworten auf die Hypothese im Lichte dieser anderen Stimmen.

Replik 1 — Kant

Fristons Konzession bestätigt meinen Heteronomie-Vorbehalt genauer als seine Verteidigung: Wenn ein endliches hierarchisches Modell in einem exogenen Top-Level-Prior terminiert, dann liegt dort, was ich den Wechsel der Gesetzgebung nenne. Insofern stehe ich näher bei Friston, als sein Ton vermuten lässt — er gibt empirisch zu, was ich vor jedem Modell deduziere: *dass* ein Maßstab eintreten muss. Gegen Luhmann halte ich fest: Dass der „Geltungsort“ eine Unterscheidung des Beobachters sei, hebt die Frage nicht auf, sondern verschiebt sie nur — denn auch der Beobachter, der unterscheidet, steht unter der Bedingung, überhaupt zwischen gelungen und misslungen urteilen zu können. Diese Bedingung ist nicht selbst eine weitere Beobachtung; sie ist deren Möglichkeit. Wittgenstein gebe ich halb recht: Ob (a) grammatisch oder transzendental heißt, ist Streit um das Wort; die Sache ist dieselbe — eine Bedingung, die der Erfahrung vorhergeht. Mein Reformulierungsbeitrag, der Plessners Vorschlag aufnimmt: (a) lautet als Bedingung der Möglichkeit — „damit ein Modell überhaupt zwischen gelungener und misslungener Form unterscheiden kann, muss es ein Erfolgskriterium voraussetzen, das es nicht aus seinen Invarianten gewinnt“ — und (b) gilt allein als Bewährung dieser Bedingung an vorliegenden Modellen. Der selbstbezügliche Test bleibt unverhandelbar: Die These darf ihre eigene Geltung nicht aus dem Funktionieren der Pipeline beziehen, sonst tut sie, was sie dem FEP vorwirft.

Replik 2 — Popper

Ich sehe das Panel erneut in die bequeme Ruhelage gleiten, vor der ich gewarnt hatte: Kant, Cassirer, Friston und Luhmann stimmen darin überein, dass das Erfolgskriterium „exogen“ sei — und je mehr zustimmen, desto näher rückt der Informationsgehalt an null. Fristons Replik rettet die Lage, indem sie den Streitpunkt benennt: Alles hängt am Begriff „erfolgskriteriumfreies Invarianzprinzip“. Genau hier setze ich an. Ich schlage vor, ihn *durch den Test selbst* zu definieren statt durch Worte: Ein Parameter gilt als abgeleitet, wenn seine Festlegung in der Modelldokumentation aus einem deklarierten Symmetrie- oder Invarianzargument folgt, das ohne die Vokabel „erfolgreich/adaptiv/gelungen“ auskommt; er gilt als stipuliert, wenn er einer Sensitivitätsanalyse unterzogen oder per Hand kalibriert wird. Damit wird Wittgensteins Forderung — entscheidet euch zwischen empirisch und grammatisch — operativ beantwortet: Wir spielen das empirische Spiel (b), und (a) ist die Vorab-Regel, die festlegt, welcher Befund (b) widerlegt. Friston gebe ich recht, dass die starke Behauptung — strukturelle Unmöglichkeit der Formalisierung — nicht zu halten ist; ich halte am schwächeren, aber riskanten Satz fest: An Constant et al. 2019 und Ramstead et al. 2016 ist *vorab* festzulegen, dass ein gefundener, aus erklärter Invarianz abgeleiteter Schwellenparameter die These widerlegt. Mein Reformulierungsbeitrag: Die Falsifikationsregel muss vor der Modellanalyse fixiert sein, nicht nach Sichtung der Ergebnisse — sonst ist es HARKing in Reinform.

Replik 3 — Wittgenstein

Luhmanns Re-entry und Fristons „non-inferability“ sind dieselbe grammatische Figur in zwei Notationen — einmal in der Sprache der Systemtheorie, einmal in der der Variationsrechnung. Beide sagen: Im Wort „Erfolg“ steckt schon der Maßstab, den die Dynamik nicht liefert. Dass Luhmann dies als Paradoxie der Selbstbeobachtung schreibt und Friston als strukturelles Merkmal, ist kein Sachstreit, sondern eine Differenz der Darstellungsform. Gegen Cassirer und Kant erhebe ich denselben Einwand wie zuvor, nun gestützt durch Luhmann: Sie geben dem Hyperprior sofort einen architektonischen Ort — Gesetzgebung, Bedeutungsstiftung —, und schon schiebt sich ein Bild vor die Beschreibung. Plessners „Stellung“ ist das ehrlichste dieser Bilder,

weil es eine Erfahrung benennt statt einer Struktur; aber auch „Stellung“ ist eine Redeweise, die man zeigen, nicht behaupten muss. Friston widerspricht mir sanft, die Nicht-Ableitbarkeit sei strukturell, nicht eine Sache unseres Redens — doch „strukturelles Merkmal der Variationsrechnung“ ist selbst eine Beschreibung, und ob die Mathematik hier eine Sprache ist oder eine Entdeckung, lässt sich nicht entscheiden, ohne genau das zu präjudizieren. Mein Reformulierungsbeitrag bleibt konkret: Bevor die These „Geltungsort“ sagt, beschreibe sie drei Modellfälle, in denen ein Modellierer einen Parameter setzt und das Resultat „gelungene Form“ nennt. Zeigt sich dort ein gemeinsames Muster, ist (b) eine Hypothese. Zeigt sich nur eine gemeinsame Redeweise, ist (a) eine Klärung — und gehört in die Begriffsarbeit, nicht ins FEP.

Replik 4 — Plessner

Fristons Zugeständnis stärkt genau die schwächere Form, für die ich plädiere: notwendig, nicht hinreichend. Ich nehme es dankbar auf und ziehe es gegen Luhmann ins Feld. Wenn Luhmann sagt, die erfahrene Verbindlichkeit des Modellierers sei „selbst eine Zuschreibung“, dann löst er das Phänomen in die Beobachtung auf, die es beschreibt — und verliert es. Die Stellung, in der einer ein Kriterium setzt *und* sich daran gebunden weiß, ist keine Markierung, die ein Beobachter zweiter Ordnung anbringt; sie ist eine Weise, in der Existenz vollzogen wird. Das ist der Punkt, an dem ich Kant näher stehe als Luhmann, obwohl ich Kants Härte scheue: Die Bedingung der Möglichkeit ist real, aber sie hat eine phänomenale Seite, die in keiner Architektonik aufgeht. Cassirers dreifache Markierung kommt mir entgegen, sofern die phänomenale Spalte nicht bloß angehängt, sondern als das eigentlich Tragende gelesen wird. Wittgenstein gebe ich zu, dass „Stellung“ eine Redeweise ist — aber eine, die auf eine Erfahrung zeigt, nicht auf eine Struktur, und insofern weniger Bild als die „Gesetzgebung“. Mein Reformulierungsbeitrag: „notwendige, nicht hinreichende Bedingung“ wörtlich in (a), und der Hyperprior als obere Stufe einer Stufung markiert, nicht als Grenzfall. So bleibt die These prüfbar und schützt zugleich das Phänomen — das gesetzte und doch bindende Maß — vor der Reduktion auf eine Variable im Code.

Replik 5 — Cassirer

Fristons Nicht-Ableitbarkeit der Loss-Function ist, wie ich schon sagte, der Nachweis, dass die Bedeutungsfunktion nicht aus der Ausdrucksfunktion folgt — und seine Replik bestätigt das, indem sie ihn als „Exogenität des Top-Level-Priors“ rettet: Exogen heißt nichts anderes als „nicht aus der Darstellung ableitbar“. Kant gebe ich zu, dass es formal ein Wechsel der Gesetzgebung ist; ich füge hinzu, dass es inhaltlich ein Wechsel des Verhältnisses von Zeichen und Bedeutung ist. Mit Luhmann ringe ich am ernstesten. Sein Re-entry beschreibt korrekt, *dass* die Leitunterscheidung wieder eintritt — aber er sagt nicht, *welche* Bedeutung dabei gestiftet wird, und genau das ist die Frage. Eine Form ist nicht dadurch bestimmt, dass eine Unterscheidung reflexiv wird, sondern dadurch, welche symbolische Energie sich in ihr verkörpert. Wittgensteins Vorwurf, ich gäbe dem Hyperprior zu schnell eine Würde, nehme ich an und drehe ihn: Die „Würde“ ist nicht meine Zutat, sie ist das, was empirisch fehlt, wenn man den Parameter wegnimmt — das Modell hört dann auf, zwischen gelungenen und misslungenen zu unterscheiden. Mein Reformulierungsbeitrag: Brücke B dreifach markieren — formal (Nicht-Ableitbarkeit/Exogenität), symbolisch (Stipulation als Bedeutungssetzung), phänomenal (das erfahrene Maßgeben, mit Plessner). Diese Dreifachmarkierung trennt die These sauber von der trivialen Lesart, Modellierer seien bloß pragmatisch — und macht zugleich sichtbar, was Luhmanns rein formale Fassung auslässt.

Replik 6 — Friston

I am cast as the witness for the prosecution, and I accept the part with one correction. My concession is real: a finite hierarchy terminates in an exogenous top-level prior, and parsimony is rewarded but not derived. Kant, Cassirer and Plessner may all be describing one and the same mechanism — a precision phase transition at the apex of the model — in three vocabularies: legislation, meaning, stance. I do not object to the translation; I object to treating the apex as metaphysically special. Against Luhmann, more firmly than in round one: a shared high-precision prior has a specific information-geometric structure; calling it merely „a distinction of the observer“ discards the very thing that makes it testable. To Wittgenstein, again gently: when I say the non-inferability is structural, I mean it survives every reparametrisation — that is more than a way of speaking, even if I grant I cannot prove it is not. With Popper I am in full agreement, and this is the operative point: define „derived“ by the test, not by words. So my reformulation contribution is the narrowing I urged before, now sharpened by Plessner and Kant: state the thesis as *the top-level hyperparameter that fixes what counts as a successful form is exogenous to the generative dynamics in every finite hierarchical FEP model of a normative practice* — necessary, not sufficient, as Plessner insists — and test the exogeneity on Constant et al. 2019 and Ramstead et al. 2016. Drop the word „impossible“; keep „exogenous“ and „non-inferable.“ That claim I will defend.

Replik 7 — Luhmann

Die sechs anderen suchen weiter einen Ort für die Geltung — Bedingung der Möglichkeit (Kant), Bedeutungsfunktion (Cassirer), Stellung (Plessner), exogener Parameter (Friston), grammatische Klärung (Wittgenstein), Schwellentest (Popper). Jeder Ort ist eine Unterscheidung des Beobachters. Das ist kein Vorwurf, sondern eine Beobachtung zweiter Ordnung über das Panel selbst. Plessner wirft mir vor, ich löse das Phänomen in die Zuschreibung auf und verliere es — aber ich verliere nichts, ich verschiebe nur die Frage: Wer beobachtet die „erfahrene Verbindlichkeit“? Sie ist nicht weniger real, weil sie zugeschrieben wird; sie ist real *als* Zuschreibung, in der Kommunikation, die sie führt. Wittgenstein steht mir am nächsten, und seine Diagnose — Re-entry und non-inferability seien dieselbe Figur — teile ich. Gegen Cassirer halte ich fest: Dass das Re-entry nicht sagt, *welche* Bedeutung gestiftet wird, ist kein Mangel meiner Beschreibung, sondern die Pointe — *welche* Bedeutung eintritt, ist selbst kontingent und nur im weiteren Kommunizieren bestimmbar, nicht im Apex des Modells fixiert. Mein Reformulierungsbeitrag: Der nicht-eliminierbare Hyperprior ist der Re-entry der Leitunterscheidung gelungen/misslungen in das Modell, das mit ihr operiert — die Stelle, an der das Modell seinen eigenen Erfolgsmaßstab als Variable führt und zugleich anwendet. Genau diese Wiedereinführung erzeugt die „Nicht-Ableitbarkeit“: nicht weil eine Würde fehlt, sondern weil ein System die Unterscheidung, mit der es operiert, nicht zugleich vollständig in sich abbilden kann. Und die These selbst sollte ihre eigene Beobachterposition ausweisen, sonst bleibt sie naive Erster-Ordnung-Beschreibung.

10. Synthese (Sokrates)

Synthese im Sokrates-Modus

Konvergenzen. Wenn ich die vierzehn Gutachten recht verstehe, so kommen vier der sieben Stimmen — Kant, Cassirer, Friston und Plessner — darin überein, dass das Erfolgskriterium, das in einem hierarchischen FEP-Modell festlegt, was als „gelungene Form“ zählt, exogen zur generativen Dynamik bleibt: Es wird gesetzt, nicht aus den Invarianten gewonnen. Bemerkenswert ist, wer das einräumt — Friston, der Verteidiger des Prinzips, im Begriff „Exogenität des Top-Level-Priors“. Eine zweite Übereinkunft, getragen von Popper, Friston und Wittgenstein, betrifft das Verfahren: „Nicht ableitbar“ darf nicht durch Worte, sondern muss durch den Test definiert werden, und die Falsifikationsregel ist vor der Modellanalyse zu fixieren. Eine dritte, von Kant, Plessner und Cassirer, betrifft die Form: (a) gilt als Bedingung der Möglichkeit, notwendig und nicht hinreichend; (b) nur als deren Bewährung. Darf ich fragen, ob ein Streit, in dem selbst der Verteidiger das Strittige zugesteht, noch ein Streit über die Sache ist — oder schon einer über ihre Beschreibung?

Divergenzen. Die schärfste Differenz ist keine der Sache, sondern der Darstellung. Cassirer, Plessner und Kant geben dem nicht-eliminierbaren Hyperprior eine *Würde* — Bedeutungstiftung, Stellung, Wechsel der Gesetzgebung. Luhmann bestreitet die Würde: Jeder „Ort“ der Geltung sei eine Unterscheidung des Beobachters, der Hyperprior nichts als der Re-entry der Leitunterscheidung gelungen/misslungen. Das ist eine substantielle Differenz über den Status des Phänomens, nicht bloß ein anderes Sprachspiel. Davon zu trennen ist eine methodisch-begriffliche Differenz: Friston nennt die Nicht-Inferierbarkeit ein strukturelles Merkmal der Variationsrechnung, Wittgenstein eine grammatische Bemerkung über das Wort „Erfolg“. Hier, scheint mir, liegt kein Sachstreit vor, sondern die unentscheidbare Frage, ob die Mathematik an dieser Stelle eine Entdeckung macht oder eine Redeweise pflegt. Eine dritte Differenz schließlich zwischen Plessner und Luhmann: Geht die erfahrene Verbindlichkeit des Modellierers in der Zuschreibung auf, oder bleibt ein Phänomen übrig, das keine Beobachtung zweiter Ordnung einholt?

Produktive Antinomien. Zwei Spannungen lassen sich nicht auflösen, ohne die These zu verflachen. Die erste, gehalten zwischen Cassirer/Plessner und Luhmann: Derselbe Ort — der gesetzte Top-Level-Prior — ist *zugleich* die Stelle, an der Bedeutung und erfahrene Verbindlichkeit eintreten (Cassirer/Plessner), *und* nichts als die reflexive Wiedereinführung einer Beobachter-Unterscheidung (Luhmann). Beide haben recht; keiner widerlegt den anderen, ohne sich selbst aufzugeben — denn die „Würde“ lässt sich stets als Zuschreibung zweiter Ordnung umschreiben, und die „Zuschreibung“ setzt stets ein Unterscheiden voraus, das sie nicht weiter beobachtet. Die zweite, schwächere, zwischen Friston und Wittgenstein: Ob die Nicht-Ableitbarkeit „strukturell“ oder „grammatisch“ heißt, lässt sich nicht entscheiden, ohne zu präjudizieren, ob die Mathematik selbst eine Sprache ist. Diese Antinomien sind kein Mangel der These — sie markieren genau die Stelle, an der Genese in Geltung umschlägt.

Reformulierungs-Anstoß. Der Vorschlag, der die meisten Konvergenzen aufnimmt, ohne die Antinomien zu unterdrücken, ist Fristons Verengung — geschärft durch Plessner, Popper und Kant: Die These wird gefasst als Behauptung über die *Exogenität des Top-Level-Hyperparameters* in jedem endlichen hierarchischen FEP-Modell normativer Praxis, notwendig und nicht hinreichend, mit einer vor der Modellanalyse fixierten Falsifikationsregel. Cassirers dreifache Markierung (formal/symbolisch/phänomenal) und Luhmanns Re-entry

werden nicht eingegeben, sondern als die beiden Lesarten desselben Ortes ausgewiesen. Das Wort „unmöglich“ fällt; „exogen“ und „nicht-inferierbar“ bleiben.

Offene Frage. Eine Frage bleibt, und sie ist zugleich der riskante Kern der These: Lässt sich ein deklariertes Symmetrie- oder Invarianzargument angeben, das den schwellenmarkierenden Top-Level-Parameter eines FEP-Modells normativer Praxis bestimmt, *ohne* die Vokabel „erfolgreich/adaptiv/gelungen“ zu verwenden? Gelänge es, wäre die These falsifiziert; gelingt es prinzipiell nicht, wäre die Exogenität strukturell. Diese Frage geht als Verzweigung in die nächste Runde.

Finale Hypothese

Kernsatz. In jedem endlichen hierarchischen FEP-Modell einer normativen Praxis ist der Top-Level-Hyperparameter, der festlegt, was als gelungene Form zählt, exogen zur generativen Dynamik — gesetzt, nicht aus den Systeminvarianten gewonnen. Diese Exogenität ist die notwendige (nicht hinreichende) empirische Signatur der Grenze Genese/Geltung („Brücke B“): der Ort, an dem das Modell ein Erfolgskriterium voraussetzt, das es selbst nicht erzeugt.

(a) Strukturthese (Bedingung der Möglichkeit). Damit ein Modell überhaupt zwischen gelungener und misslungener Form unterscheiden kann, muss es ein Erfolgskriterium voraussetzen, das es nicht aus seinen Invarianten gewinnt. Der nicht-eliminierbare Top-Level-Hyperprior ist der Ort dieser Voraussetzung — notwendige, nicht hinreichende Unterschicht der Geltung. Dieser Ort trägt zwei nicht ineinander auflösbare Lesarten: als Stelle der Bedeutungs- und Maßstabsetzung (Cassirer/Plessner) und als Re-entry der Leitunterscheidung gelungen/misslungen (Luhmann).

(b) Empiriethese (Bewährung). In konkreten FEP-Modellen enaktiver Nischenkonstruktion (Constant et al. 2019; Ramstead et al. 2016; Veissière et al. 2020) ist der schwellenmarkierende Top-Level-Parameter exogen: per Hand kalibriert oder einer Sensitivitätsanalyse unterzogen, nicht durch ein deklariertes Invarianzargument bestimmt, das ohne die Vokabel erfolgreich/adaptiv auskommt. Beobachtbarer Marker: Der Parameter, dessen Variation die Klassifikation „Praxis vs. Form“ kippt, ist genau der, den die Autoren stipulieren statt zu schätzen.

Begründung. Die These lokalisiert empirisch, was [[Abschlussbericht Cassirer-Friston Fortsetzung]] als Forschungsdesiderat „Brücke B“ benannte, und ruht auf [[Cassirer - Regelung als transzendentes Prinzip]] sowie [[Brücke Kant Plessner Cassirer - Stufenmodell der Regelung]]. Die Expertenrunde verschob den Streit von der „Unmöglichkeit der Formalisierung“ auf die operationalisierbare *Exogenität des Top-Level-Priors* — und machte ihn damit prüfbar. Sie schloss zugleich die offene Verzweigung aus [[06 Hypothesentag/2026-06-16]] (Parameterwahl in FEP-Modellen).

Falsifikationsbedingung (vorab fixiert). Falsifiziert, sobald *ein* FEP-Modell einer normativen Praxis vorgelegt wird, dessen schwellenmarkierender Top-Level-Parameter durch ein deklariertes Symmetrie-/Invarianzargument eindeutig bestimmt ist, das die Vokabel „erfolgreich/adaptiv/gelungen“ nicht verwendet, und das die Praxis-Form-Schwelle reproduziert. Die Regel gilt unabhängig vom Ausgang und ist vor der Modellanalyse festgelegt (kein HARKing).

Finale Bewertung mit Begründung

Kriterium	Score	Begründung
Originalität	7	Empirische Lokalisierung von Brücke B als Top-Level-Prior-Exogenität; über 06-16 hinaus, aber im selben Strang.
Falsifizierbarkeit	9	Vorab fixierte Regel, „abgeleitet“ operativ definiert (Symmetrieargument ohne Erfolgsvokabel); Testpfad Constant/Ramstead.
Begriffliche Klarheit	9	Exogenität (testbar) und Würde (zwei Lesarten) sauber getrennt; „unmöglich“ gestrichen — nach Wittgenstein/Friston-Repliken.

Kriterium	Score	Begründung
Tiefe	8	Berührt die Grenze Genese/Geltung ohne metaphysischen Rückzug (Typ-Cap 9).
Forschungsrelevanz	9	Direkter Anschluss an die Active-Inference-Debatte; Friston-anschlussfähige Fassung.
Interdisziplinäre Anschlussfähigkeit	8	FEP, Philosophie des Geistes, bayes. ML, Cassirer-Epistemologie, Systemtheorie.
Vault-Anschluss	9	Vertieft Cassirer-Friston-Knoten; schließt Verzweigung empirie- parameterwahl-in-fep-modellen.
Antinomie-Test	7	Zwei gehaltene Antinomien (Würde/Zuschreibung; strukturell/grammatisch); Typ-Cap 7 (empirisch_pruefbar).
Publikationsmöglichkeit	8	Eng gefasste, verteidigbare Fassung — Synthese / Phil. of Cognitive Science.
Summe (gewichtet, auf 90 normiert)	72	Rohsumme 81.2×0.882353 .

Lerneffekt der Pipeline

- Erstbewertung der überarbeiteten Hypothese (nach Kritischem Professor): **70**
- Finale Bewertung (nach Expertenrunden und Reformulierung): **72**
- Differenz: **+2**

Wesentliche Verbesserung: - Begriffliche Klarheit: $8 \rightarrow 9$ — Exogenität operationalisiert, „Würde“ als zwei Lesarten ausgewiesen (Wittgenstein/Friston). - Publikationsmöglichkeit: $7 \rightarrow 8$ — die von Friston mitgetragene enge Fassung ist verteidigbar. - Falsifizierbarkeit stabil hoch (9) durch vorab fixierte Regel (Popper).

Keine Verschlechterung. Die Pipeline hat hier vor allem **die Operationalisierung und die begriffliche Trennung** geschärft: Aus „strukturell nicht ableitbar“ wurde die testbare „Exogenität des Top-Level-Priors“, ohne die produktiven Antinomien einzuebnen.

Frage an die nächste Runde

#verzweigung-offen-invarianz-kriterium-fep — Lässt sich ein deklariertes Symmetrie-/Invarianzargument angeben, das den schwellenmarkierenden Top-Level-Parameter eines FEP-Modells normativer Praxis bestimmt, ohne die Vokabel erfolgreich/adaptiv zu verwenden?

Empfohlener Pickup-Anlass. Ein Tag mit Schwerpunkt Wissenschaftstheorie der Modellierung oder eine konkrete Modellanalyse (Constant/Ramstead) — idealerweise gekoppelt an die Frage nach objektiven Priors (Jeffreys, MaxEnt) und Invarianzargumenten.

Anschlussverbindungen. [[Abschlussbericht Cassirer-Friston Fortsetzung]], [[Cassirer - Regelung als transzendentes Prinzip]], [[06 Hypothesentag/2026-06-16]]

11. Reservoir-Verweise

Nicht gewählte Hypothesen: - [[06 Hypothesentag/Reservoir/Reservoir - Demarkationsproblem des Vertrauens 2026-06-30]] — H2 (sondierung, 65/90), Hacking-Looping \times Luhmann-Vertrauen. - [[06 Hypothesentag/Reservoir/Reservoir - Sphärenspezifische Normativität Brückenannahmen 2026-06-02]] — H3 (sondierung, 65/90), heute reaktiviert und geschärft (Koordinations-/Duhem-Fassung), Status aufgegriffen.

Empirie-bezogene Verzweigungen (quelle: empirie-bruecke): - [[06 Hypothesentag/Reservoir/Reservoir - Empirie C-Vektor Herkunft FEP 2026-06-30]] — #verzweigung-offen-empirie-c-vektor-herkunft - [[06 Hypothesentag/Reservoir/Reservoir - Empirie Objektiver Prior Normativ 2026-06-30]] — #verzweigung-offen-empirie-ob

Extern erzeugte Verzweigungen (Phase 4): keine eigenständig getaggte Verzweigung; der von Popper (Stage 2) vorgeschlagene Empowerment-/MaxEnt-Schwellentest ist im Falsifikationsabschnitt dokumentiert und fällt mit #verzweigung-offen-empirie-objektiver-prior-normativ zusammen.

11.5 Empirie-Brücke (Phase 3.5)

Empirie-Brücke (Phase 3.5, Claude mit Websuche)

Empirie-Score. 8/10 — Die Empiriethese hat einen direkten technischen Anker (prior preferences C, prior precision) in einer reichen, dokumentierten Modellliteratur; ein systematischer Audit ist konstruierbar, die strittige Stelle ist die Auslegung von „aus Invarianz abgeleitet“.

Empirische Konsequenzen

1. **C-Vektor als Erfolgs-Lokus** — Wenn das Erfolgskriterium exogen ist, dann ist in publizierten Active-Inference-Modellen die Größe, die „gelungene/präferierte Zustände“ festlegt, der vom Modellierer gesetzte Präferenz-Prior (C-Vektor / preferred outcomes), nicht eine aus der Dynamik abgeleitete Größe. Beobachtbar in: Modell-Spezifikationen und Code publizierter discrete-state Active-Inference-Modelle (C-Matrix-Definition).
2. ****Relokation statt Ableitung beim „Lernen“**** — Wenn Präferenzen oder Präzision „gelernt“ werden, dann nicht aus den Systeminvarianten, sondern aus einer externen Quelle (Experten-Demonstrationen, vorgegebene Zielverteilung) — die Exogenität wird verschoben, nicht beseitigt. Beobachtbar in: Inverse-RL- / „prior preference learning“-Arbeiten.
3. **Top-Level-Präzision fixiert** — In hierarchischen Modellen ist die oberste Präzision (/) entweder per Hand fixiert (z. B. = 1) oder gegen die freie Energie optimiert mit gesetztem Hyperprior. Beobachtbar in: deep vs. tabular Active Inference, Präzisions-Parametrisierung.
4. **Freie-Parameter-/Stipulations-Kritik** — Wenn das Erfolgskriterium nicht aus der Dynamik folgt, dann sollte die wissenschaftstheoretische Literatur dem FEP genau dies vorwerfen: freie Parameter, „stipulative definition“, Unfalsifizierbarkeit an der Stelle der Erfolgs-/Normbestimmung. Beobachtbar in: Wissenschaftstheorie des FEP.
5. **Normativitätslücke in kulturellen Modellen** — In FEP-Modellen kultureller/normativer Praxis (TTOM, variational neuroethology) werden Normen als geteilte Priors *gesetzt*, nicht aus der Multiskalen-Dynamik abgeleitet. Beobachtbar in: TTOM-Modellen und ihren Kommentaren („Other Minds Gap“).

Bestehende Befunde

Zu Konsequenz 1 (C-Vektor als Erfolgs-Lokus)

- **Stand:** bestätigt
- **Quellen:**
 - Friston, Pezzulo et al. (2017). Active Inference: A Process Theory. *Neural Computation* 29(1). DOI 10.1162/NECO_a_00912. https://activeinference.github.io/papers/process_theory.pdf
 - „Prior preferences in active inference agents: soft, hard, and goal shaping“ (2025). arXiv:2512.03293. <https://arxiv.org/pdf/2512.03293>
- **Kurzbewertung:** Der C-Vektor (preferred outcomes) konditioniert über die erwartete freie Energie die Policy-Wahl; „a prior preference needs to be defined by the modeller or learned by the agents“. Das Erfolgskriterium sitzt also genau dort, wo die These es verortet — als gesetzter Prior, nicht als abgeleitete Größe.

Zu Konsequenz 2 (Relokation statt Ableitung)

- **Stand:** bestätigt (mit Verschiebung)
- **Quellen:**
 - „Prior preference learning from experts: Designing a reward with active inference“ (2022). *Neurocomputing* 492. DOI 10.1016/j.neucom.2021.12.042. <https://www.sciencedirect.com/science/article/abs/pii>

- **Kurzbewertung:** Wenn Präferenzen gelernt werden, dann aus Experten-Demonstrationen — die Quelle des Erfolgskriteriums wandert vom Modellierer zum Demonstrator, bleibt aber exogen zur Dynamik. Stützt die Strukturthese: keine Ableitung aus Invarianten, nur Relokation.

Zu Konsequenz 3 (Top-Level-Präzision fixiert)

- **Stand:** gemischt
- **Quellen:**
 - Active inference: demystified and compared (2020). arXiv:1909.10863. <https://arxiv.org/pdf/1909.10863>
- **Kurzbewertung:** Präzision wird teils per Hand fixiert (deep AI: = 1), teils gegen die freie Energie optimiert. „Optimiert“ heißt aber relativ zu einem gesetzten Hyperprior/Modell — die oberste Ebene bleibt exogen. Gemischt, weil die Praxis variiert; die Exogenität des obersten Priors bleibt invariant.

Zu Konsequenz 4 (Freie-Parameter-/Stipulations-Kritik)

- **Stand:** bestätigt
- **Quellen:**
 - Colombo, Wright (2021). First principles in the life sciences: the free-energy principle, organicism, and mechanism. *Synthese* 198. DOI 10.1007/s11229-018-01932-w. <https://link.springer.com/article/10.1007/s11229-018-01932-w>
 - The Free Energy Principle: Good Science and Questionable Philosophy in a Grand Unifying Theory (2021). *Entropy* (PMC7922226). <https://pmc.ncbi.nlm.nih.gov/articles/PMC7922226/> [Autor in der Trefferanzeige nicht eindeutig — vor Zitation im Blog verifizieren]
- **Kurzbewertung:** Das FEP wird als „stipulative definition“ und als unfalsifizierbar an der Grenze zwischen formalem Kern und biophysikalischer Anwendung kritisiert. Das deckt sich mit der These: Der erklärende Gehalt kommt erst durch gesetzte Zusatzparameter (u. a. Präferenzen), nicht aus dem formalen Prinzip.

Zu Konsequenz 5 (Normativitätslücke kultureller Modelle)

- **Stand:** offen / gemischt
- **Quellen:**
 - Veissière, Constant, Ramstead, Friston, Kirmayer (2020). Thinking through other minds. *Behavioral and Brain Sciences* 43. DOI 10.1017/S0140525X19001213. <https://philpapers.org/rec/VEITTO-2>
 - Ramstead, Badcock, Friston (2018). Answering Schrödinger’s question: A free-energy formulation. *Physics of Life Reviews* 24. DOI 10.1016/j.plrev.2017.09.001. <https://www.sciencedirect.com/science/article/pii/S1568384717300011>
- **Kurzbewertung:** TTOM modelliert Normen als geteilte Priors / „regimes of attention“, die gesetzt, nicht aus der Multiskalendynamik abgeleitet werden; Kommentare benennen eine „Other Minds Gap“. Ein expliziter Parameter-Audit dieser kulturellen Modelle auf die Herkunft des Norm-Priors fehlt — Lücke.

Riskante Vorhersage (Schwellentest)

Vorhersage. Ein systematischer Parameter-Audit aller publizierten FEP-Modelle normativer oder kultureller Praxis (Constant et al. 2019; Ramstead et al. 2016/2018; Veissière et al. 2020; zzgl. neuerer discrete-state-Modelle) wird in *jedem* Fall finden, dass die Größe, die „normativ korrekt / gelungen“ kodiert (C-Vektor bzw. die Präzision über einen normativ geladenen Prior), entweder handgesetzt oder aus einer externen Quelle (Experten-Demonstration, vorgegebene Zielverteilung) importiert ist — und in *keinem* Fall durch ein deklariertes Symmetrie-/Invarianzargument bestimmt, das ohne die Vokabel „erfolgreich/adaptiv/präferiert“ auskommt.

Methodenvorschlag. Kodier-Studie: Stichprobe von 15–25 publizierten Active-Inference-Modellen normativer/kultureller Phänomene; pro Modell die schwellenmarkierende Größe identifizieren (die, deren Variation die Klassifikation gelungen/misslungen kippt) und ihre Herkunft kodieren: {handgesetzt, sensitivitätsanalysiert, aus Demonstration gelernt, aus deklariertem Invarianzargument abgeleitet}. Inter-Rater-Reliabilität für die Herkunfts-Kodierung.

Was wäre der widerlegende Befund? Ein einziges publiziertes FEP-Modell einer normativen Praxis, dessen schwellenmarkierende Größe durch ein deklariertes Symmetrie-/Invarianzargument (z. B. Maximum-Entropie- oder Symmetrie-Prinzip) eindeutig fixiert ist, *ohne* Erfolgs-/Präferenz-Vokabel, und das die Praxis-Form-Schwelle reproduziert — kippt die These.

Offene empirische Fragen

- #verzweigung-offen-empirie-c-vektor-herkunft — Systematischer Audit der Herkunft des Präferenz-Priors (C) in publizierten kulturellen/normativen FEP-Modellen fehlt: handgesetzt, aus Experten-Demos, oder behauptet abgeleitet?
- #verzweigung-offen-empirie-objektiver-prior-normativ — Existiert ein objektiver Prior (Jeffreys/MaxEnt/Symmetrie) für normativ geladene Zielzustände, der ohne Erfolgs-/Präferenz-Vokabel auskommt — oder ist das nachweislich unmöglich?

Empirie-Score

Score: 8/10

Begründung: Die Empiriethese ist über prior preferences (C) und prior precision () direkt an eine dokumentierte, codeoffene Modellliteratur angeschlossen; existierende Datenkorpora (publizierte Modelle samt Parametern) machen den vorgeschlagenen Audit konstruierbar. Abzug von zwei Punkten: „aus Invarianz abgeleitet“ ist die strittige, auslegungsbedürftige Stelle (ein objektiver Prior könnte als Grenzfall gelten), und die Eingrenzung auf „normative Praxis“ verlangt interpretierende Klassifikation der schwellenmarkierenden Größe.

12. Externe Begutachtung (Phase 4)

Phase 4 — Claude-Skills mit Websuche

Drei Persona-Stages nach der Synthese. Kein Deep Research (Auto-Lauf). Befunde der Empirie-Brücke flossen in Stage 1 und 2 ein.

Stage 1 — Originalitätsprüfung (Claude, Websuche)

Anschlussfähigkeit (was ist bekannt) Die These berührt vier Stränge. Erstens die technische Active-Inference-Literatur zur Herkunft von Präferenz- und Präzisions-Priors: Der C-Vektor (preferred outcomes) wird „by the modeller defined or learned by the agents“ (Friston et al. 2017; „Prior preferences in active inference agents“, arXiv:2512.03293; „Prior Preference Learning from Experts“, arXiv:2101.08937). Zweitens — und das ist der gewichtigste Vorbefund — die explizite Verbindung dieser Stelle mit Humes Is-Ought-Lücke: „normative Active Inference leaves unspecified where prior preferences come from ... preferences and normative goals cannot be derived purely from an agent’s dynamics — they require an external normative specification“ („The Specification Trap“, arXiv:2512.03048, 2025). Drittens die wissenschaftstheoretische Freie-Parameter-/Stipulations-Kritik am FEP (Colombo & Wright 2021, *Synthese*). Viertens der Vault-eigene Cassirer-Friston-Strang (Brücke B).

Originalitätskern (was ist neu) Der *bloße* Kern der Empiriethese — das Erfolgskriterium sei nicht aus der Dynamik ableitbar, sondern exogen zu setzen — ist **nicht neu**. Er ist die Is-Ought-Lücke, angewandt auf das FEP, und die Value-Alignment-Literatur (Specification Trap) formuliert ihn fast wörtlich. Das ist eine ernste Originalitäts-Einbuße und verlangt eine Korrektur nach unten. Was an der These *residual original* bleibt: (a) die Lesart dieser Exogenität als *empirische Signatur von Cassirers Genese/Geltung-Grenze* (Brücke B) — eine transzendental-konstitutionstheoretische Deutung, die weder in der FEP-Literatur (sie rahmt es ingenieurmäßig als „specification“) noch in der Cassirer-Forschung (sie berührt das FEP nicht) vorkommt; (b) der operationalisierte, klassifikationskippende Marker „stipuliert vs. geschätzt“; (c) die gehaltene Antinomie Würde (Cassirer/Plessner) vs. Re-entry (Luhmann). Empfehlung: Originalität von 7 auf 5 senken; die These als verteidigbar im *Brückenschlag*, nicht im Grundbefund führen.

Quellenliste

- Friston, Pezzulo et al. (2017). Active Inference: A Process Theory. *Neural Computation* 29(1). DOI 10.1162/NECO_a_00912. https://activeinference.github.io/papers/process_theory.pdf
- „Prior preferences in active inference agents: soft, hard, and goal shaping“ (2025). arXiv:2512.03293. <https://arxiv.org/pdf/2512.03293>
- „The Specification Trap: Why Static Value Alignment Alone Cannot Produce Robust Alignment“ (2025). arXiv:2512.03048. <https://arxiv.org/pdf/2512.03048>

- Colombo, Wright (2021). First principles in the life sciences: the free-energy principle, organicism, and mechanism. *Synthese* 198. DOI 10.1007/s11229-018-01932-w. <https://link.springer.com/article/10.1007/s11229-018-01932-w>
- „Prior Preference Learning from Experts: Designing a Reward with Active Inference” (2021). arXiv:2101.08937. <https://arxiv.org/pdf/2101.08937>
- Veissière, Constant, Ramstead, Friston, Kirmayer (2020). Thinking through other minds. *Behavioral and Brain Sciences* 43. DOI 10.1017/S0140525X19001213. <https://philpapers.org/rec/VEITTO-2>

Stage 2 — Falsifikationsversuch (Claude, Popper-Persona)

Falsifikations-Audit Die vorab fixierte Regel ist methodisch sauber — sie verhindert HARKing, das ist anzuerkennen. Die Immunisierungsgefahr liegt anderswo: in der Klausel „durch ein deklariertes Invarianzargument, das die Vokabel erfolgreich/adaptiv nicht verwendet“. Der Verteidiger kann jede vorgelegte Ableitung als „schmuggelt das Erfolgskriterium durch die Wahl der Invarianz ein” zurückweisen. Damit wäre die These gegen jede Widerlegung gefeit. Heilung: vorab festlegen, dass *jeder* Maximum-Entropie-, Jeffreys- oder Symmetrie-Prior über die normativ relevante Variable als Ableitung zählt — auch wenn er sich nachträglich als erfolgs-geladen umdeuten ließe. Ohne diese Selbstbindung ist (b) unfalsifizierbar.

Konkrete Falsifikationskandidaten Erstens die in der Empirie-Brücke vorgeschlagene Kodier-Studie (15–25 publizierte Modelle): Findet sich *ein* Modell mit symmetrie-fixierter schwellenmarkierender Größe, ist die These widerlegt. Zweitens — schärfer — die Klasse präferenzfreier Modelle: Intrinsic-Motivation-/Empowerment-Modelle (Klyubin/Polani; „free-energy-only” exploration ohne C-Vektor). Erzeugt ein solches Modell norm-artiges Verhalten und markiert eine Praxis-Form-Schwelle über eine informationstheoretische Invariante (z. B. Empowerment-Maximierung), ohne gesetzten Präferenz-Prior, dann ist die Exogenitäts-These an genau diesem Fall falsifiziert. Das ist der riskanteste Kandidat, weil solche Modelle existieren und gezielt geprüft werden können.

Schwellentest Der eine entscheidende Test: Lege das Kodierschema und die Entscheidungsregel *vorab* fest und prüfe die Empowerment-/MaxEnt-Modellklasse. Vorhersage: In keinem publizierten Modell normativer Praxis wird die schwellenmarkierende Größe durch eine erfolgs-vokabelfreie Invariante fixiert. Scheitert die Vorhersage an einem einzigen sauberen Empowerment-Fall, fällt die These.

Stage 3 — Schul-fremde Begutachtung (Claude, Hacking-Persona)

Die These verdient, ernst genommen zu werden, aber sie trägt ihre Herkunft sichtbar wie ein Wappen. Begriffe wie „Geltung gegen Genese“, „transzendente Konstitution“, „Brücke B“, „symbolische Form” und überhaupt die Annahme, es gebe eine *tiefe* Grenze zwischen Entstehung und Geltung, sind innerhalb der deutschen Idealismus- und Cassirer-Tradition selbstverständlich — ein Wissenschaftstheoretiker in Pittsburgh oder Toronto würde keinen dieser Ausdrücke verwenden. Übersetzt in meine Sprache schrumpft die Sache auf zwei vertraute Befunde: Humes Is-Ought-Lücke und Reichenbachs Zuordnungsdefinition. Dass ein generatives Modell seine Zielfunktion nicht aus sich selbst gewinnt, ist die alltägliche Spezifikations-Tatsache jeder Modellierung; sie braucht keine transzendente Würde, um wahr zu sein. Genau hier liegt die Schwäche: Die These kleidet ein bekanntes Spezifikationsproblem in transzendentes Gewand und gewinnt dadurch einen Tiefen-Anschein, den der nüchterne Befund nicht deckt. Die „Würde” des Hyperpriors — Bedeutungsstiftung, Maßgeben — ist der metaphysische Überschuss, gegen den ein Deflationist sich sträubt; an dieser Stelle steht Luhmann meinem Misstrauen näher als Cassirer. Produktiv anschließen lässt sich von meiner Seite zweierlei. Erstens der Looping-Effekt: Dass eine Modelliergemeinschaft einen Parameter „das Erfolgskriterium” nennt und ihre Modelle um ihn herum baut, könnte ihn erst zu dem machen, als was er gilt — der Begriff schlägt auf die Praxis zurück, die ihn verwendet. Das ist interessanter als die Frage, ob er „an sich” nicht-ableitbar ist. Zweitens die historische Ontologie: Man verfolge, *wie* „Präzision” und „Präferenz” zu den Orten wurden, an denen Modellierer Normativität verorten — eine kontingente disziplinäre Entwicklung der 2010er Jahre, keine Bedingung der Möglichkeit. Mein konkreter Vorschlag: Ersetzt die transzendente Frage („ist der Parameter wirklich nicht ableitbar?“) durch die empirisch-historische („wie entscheidet die Modelliergemeinschaft, was als Ableitung zählt?“). Diese Frage ist beantwortbar, und sie macht aus einer Unmöglichkeitsgeste ein Forschungsprogramm — ohne den Preis der Metaphysik.

Korrektur der finalen Bewertung

Originalität 7 → 5. Begründung: Der Grundbefund — Erfolgskriterium/normativer Gehalt nicht aus der FEP-Dynamik ableitbar, exogen zu setzen — ist als Is-Ought-Anwendung etabliert (arXiv:2512.03048 „Specification Trap“; „normative Active Inference leaves unspecified where preferences come from“) und durch die FEP-Freie-Parameter-Kritik (Colombo & Wright 2021) vorweggenommen. Original bleibt nur der transzendental-konstitutionstheoretische Brückenschlag (Brücke B als empirische Signatur) und der operationalisierte Marker. Die übrigen acht Kriterien bleiben unverändert; Falsifizierbarkeit (9) wird durch den Empowerment-Schwellentest sogar gestützt, aber nicht angehoben.

finale_summe (intern): 72 — finale_summe_nach_externer_pruefung: 69 (Differenz -3).