

Die namenlose Fehlerspur

Vollständiges Gutachten · 12. Oktober 2026

Hauptlinie und Nebenlinie · 7 unabhängige Gutachten · 7 Kurzbegutachtungen · 7 Repliken ·
Sokrates-Synthese · Empirie und externe Prüfung

GPT-6 Astra · Reasoning max

Die Denkerstimmen sind KI-Rekonstruktionen. Die Prüfung ist ein Forschungs- und Schreibverfahren, keine Begutachtung durch die genannten Personen und kein empirischer Nachweis.

Bewertung: 63/90 vor dem Panel, 62/90 intern, 61/90 nach externer Prüfung. Empirie-Score: 7/10.

Hypothesentag — 2026-10-12

Vault-Lage und Verfahren

Der Scanner erfasste 500 Dateien, 4 offene Verzweigungen und 70 zeitfensterbezogene Virgin Nodes. Der Lauf wurde am 12.10. begonnen und über Mitternacht fortgeführt. Er behält sein ursprüngliches Wochenlauf-Datum. Alle Denkerstimmen sind KI-Rekonstruktionen. Auch die externe Prüfung wurde mit GPT-6 Astra/max durchgeführt; sie ist eine getrennte Prüfstufe mit externer Literatur, keine Begutachtung durch menschliche Fachleute oder ein unabhängiges Fremdmodell.

Verzweigungs-Pickup

Warm-Pick. empirie-ausstiegsgruende-provenienz aus 06 Hypothesentag/2026-10-05.

Cold-Pick. Genesis - Paradies als regulatorischer Nullzustand, Zieldomäne Anthropologie. Der Scanner bezeichnet die Notiz zeitfensterbezogen als virgin. Ihre tatsächliche frühere Verwendung am 20./21.05.2026 wurde verifiziert. H2 ist eine kritische Neulesung.

Drift und Echo-Chamber. Kein formaler Alarm; im Siebentagefenster liegt nur ein Bericht. Das erlaubt keine belastbare Aussage über eine wöchentliche Drift. Der Außenblick erfolgt aufgrund des ausdrücklichen Wochenauftrags.

Reservoir-Priorität. Autokatalyse als Schwellenarchitektur wurde vom Pickup angeboten, wegen Vorrangs des Außenblicks nicht konsumiert.

Außenblick

wochenschwerpunkt. Im Siebentagefenster liegt ein Bericht: epistemologisch_systemtheoretisch, mit Goethe, Carver, Luhmann und Reinach; Thema waren tradierte Ausstiegsgründe.

paradigmatische_kernannahme. Der letzte Lauf vermutete unter genau begrenzten Bedingungen einen Nutzen historisch erkennbarer Gründe für öffentlich verantwortete Revision. Der Außenblick prüft eine Gegenrichtung: Sichtbare persönliche Herkunft kann sachlich gebotene Revision erschweren. Dies widerlegt die Vorwochenhypothese nicht unmittelbar, sondern begrenzt ihre Übertragbarkeit.

gegen_these_kernsatz. In wissenschaftlichen Arbeitsgruppen kann eine nachvollziehbare Geschichte der Argumente zusammen mit vorübergehend verborgener persönlicher Urheberschaft bessere Fehlerkorrektur ermöglichen als die sofortige öffentliche Zuordnung jedes Schritts.

alternative_tradition. Experimentelle Wissenschaftssoziologie und institutioneller Fallibilismus: Verfahren der Kritik statt Autorität überlieferter Gründe.

aufhaenge_knoten. Luhmann - Vertrauen als Komplexitätsreduktion

vorgeschlagene_falsifikation. Bei gleichen Befunden und kontrollierter Informationsmenge zeigt die vorübergehend verdeckte Urheberschaft keinen Vorteil bei richtigen Korrekturen oder erhöht falsche Änderungen.

begruendung_aussenblick. Expliziter Wochenauftrag. Nur ein Bericht im Zeitfenster; deshalb keine empirisch belastbare Drift-Diagnose. Die rollierende Acht-Berichte-Basis meldet ebenfalls keinen Alarm.

Reservoir-Kandidat Autokatalyse als Schwellenarchitektur zurückgestellt, nicht konsumiert. Der Cold-Pick ist zeitfensterbezogen virgin.

Drei initiale Hypothesen

Alle drei Aussagen sind zu prüfende Kandidaten. Modell: GPT-6 Astra, Reasoning max.

H1 — Überlieferung als Einspruchsrecht

Kernsatz. Ein tradierter Ausstiegsgrund legitimiert eine Zielaufgabe nicht schon durch sein Alter, sondern indem er einen früher gewährten Ausstieg als Einwand gegen eine heutige Verweigerung verfügbar macht. Die wirksame Einheit der Überlieferung ist deshalb ein anfechtbarer Präzedenzfall, nicht ein bloß wiederholter Satz.

Begründung. Die offene Verzweigung empirie-ausstiegsgruende-provenienz aus 06 Hypothesentag/2026-10-05 fragt, wie historische Herkunft und aktuelle Anerkennung getrennt werden können. Reinach - Soziale Akte eröffnet die Unterscheidung zwischen einer inhaltlich guten Begründung und einer sozial beanspruchbaren Berechtigung. Luhmann - Vertrauen als Komplexitätsreduktion macht die Verfügbarkeit einer erwartbaren Reaktion anschlussfähig. Neu ist der Fokus auf den Einspruch gegen selektive Anwendung: Darf sich auch ein bislang nicht begünstigtes Mitglied auf den alten Fall berufen?

Falsifikationsbedingung. Eine kohärente, quellengestützte Praxis, in der ein früherer Ausstieg ausdrücklich als verbindlicher Präzedenzgrund anerkannt wird, aber keinerlei Anspruch auf Prüfung eines vergleichbaren Falls entsteht, widerlegt die behauptete Verbindung. Bloße Gewohnheit, bloßes Alter und eine einmalige Gnadenentscheidung sind separat zu prüfen.

Quelle. warm_pick: empirie-ausstiegsgruende-provenienz

Themenraum. Begriffsanalyse sozialer Berechtigung

H2 — Das Selbst im Verhör

Kernsatz. Genesis 2–3 lässt sich mit Plessners exzentrischer Positionalität als Erzählung einer nachträglichen Selbstzuschreibung lesen: Der Mensch wird nicht erst durch den Regelbruch handlungsfähig, sondern muss sein bereits vollzogenes Handeln auf eine an ihn gerichtete Frage als das eigene ausweisen.

Begründung. Genesis - Paradies als regulatorischer Nullzustand deutet die Schlange als Angebot eines anderen Selbst und den Fall als Freiheitsentstehung. Doch Gebot, Erinnerung und Antwort setzen im Erzählverlauf bereits Fähigkeiten voraus, deren Entstehung erst behauptet wird. Der anthropologische Zugriff unterscheidet deshalb Können, Sich-Verbergen und Sich-Verantworten. Der Übergang liegt möglicherweise in der umstrittenen Zuschreibung der eigenen Handlung, nicht in einer psychischen Geburt aus Bewusstlosigkeit.

Falsifikationsbedingung. Die Auslegung scheitert, wenn der Primärtext vor dem Essen keine adressierbare Handlungsfähigkeit erkennen lässt oder die spätere Befragung keinen zusätzlichen Selbstbezug gegenüber der frühen Rede erzeugt; mindestens drei Textstellen und eine konkurrierende Deutung sind zu vergleichen.

Quelle. zieldomaene: Anthropologie; cold_pick: Genesis - Paradies als regulatorischer Nullzustand

Themenraum. Hermeneutische Anthropologie der Selbstzuschreibung

H3 — Die namenlose Fehlerspur

Kernsatz. In wissenschaftlichen Arbeitsgruppen verbessert eine rekonstruierbare Geschichte von Befunden und Revisionen die sachgerechte Korrektur besonders dann, wenn persönliche Urheberchaft erst nach der gemeinsamen Entscheidung offengelegt wird. Die vorübergehende Trennung von sachlicher Herkunft und persönlicher Zuschreibung erhält das Gedächtnis des Irrtums, ohne dessen Korrektur sofort zu einer Statusentscheidung zu machen.

Begründung. Der Außenblick setzt beim Unterschied zwischen dem Grund einer Behauptung und dem Ansehen ihres Urhebers an. Luhmann - Vertrauen als Komplexitätsreduktion liefert einen Vault-Anschluss über die Entlastung durch Personenvertrauen; hier kann dieselbe Entlastung eine falsche Bindung erzeugen. Der Vorwochenbericht 06 Hypothesentag/2026-10-05 behandelt öffentlich verwendbare Ausstiegsgründe, während H3 die institutionelle Prüfung von Forschungsbehauptungen untersucht. Geprüft wird nicht pauschal der Nutzen von Anonymität, sondern ihr Zusammenspiel mit erhaltener sachlicher Revisionsgeschichte.

Falsifikationsbedingung. In einem randomisierten 2x2-Gruppenexperiment werden rekonstruierbare Befundgeschichte und Zeitpunkt persönlicher Zuschreibung unabhängig variiert. Die These scheitert, wenn bei vorhandener Geschichte verzögerte Zuschreibung richtige Korrekturen nicht verbessert, der Zusatznutzen der Geschichte nicht größer wird oder falsche Korrekturen zunehmen.

Quelle. devil_advocate

Themenraum. Wissenschaftssoziologie und kollektive Fehlerkorrektur

Gegenthese-zu. Der Übertragbarkeit des Herkunftsvorteils des letzten Wochenlaufs auf wissenschaftliche Fehlerkorrektur.

1. **H1 — Begriffsanalytisches Vorwurfs-Set.** Der Kernsatz verbindet argumentative Verfügbarkeit, einen Anspruch auf Prüfung und die Berechtigung zum Ausstieg, ohne diese Übergänge zu begründen. Ein früher gewährter Ausstieg macht einen Einwand formulierbar. Ein legitimer Anspruch verlangt zusätzlich eine normative Brücke; bloße Wiederholung reicht dafür nicht. Institutionelle Geltung und moralische Rechtfertigung bleiben unterscheidungsbedürftig.

H0/Gegendeutung: Überlieferung bewahrt argumentative Ressourcen. Ob ein Mitglied Prüfung verlangen darf, ergibt sich aus unabhängig begründeten Verfahrensnormen. Der Präzedenzfall könnte somit einen bestehenden Anspruch sichtbar machen.

Zwei Alternativen: Erstens könnte eine unabhängig gerechtfertigte Gleichbehandlungsnorm den Vergleich mit früheren Entscheidungen verlangen. Zweitens könnte eine ausdrückliche Selbstbindung einer zuständigen Instanz spätere Ansprüche begründen. Dann wäre die Verpflichtungserklärung maßgeblich; ihre Reichweite und Legitimität blieben gesondert zu prüfen.

Prüfmethode und Quellenlage: Die Falsifikationsbedingung droht zirkulär zu werden, wenn „verbindlicher Präzedenzgrund“ bereits einen individuellen Prüfungsanspruch einschließt. Als begriffliches Gegenmodell ist eine objektive Bindung der Entscheidungsinstanz ohne entsprechendes subjektives Verfahrensrecht zu untersuchen. Ein kohärentes Gegenmodell könnte den behaupteten notwendigen Zusammenhang auch ohne historische Fundstelle treffen. Praxisbeispiele können Begriffsgebrauch belegen, für sich aber keine normative Notwendigkeit beweisen. Originalpassagen aus Reinach, Luhmann und den angeführten Vault-Einträgen wurden nicht vorgelegt. Ihre argumentative Tragfähigkeit bleibt deshalb unbestätigt.

Zeitliche Logik: Trenne Erstentscheidung, spätere Überlieferung und heutige Anspruchserhebung. Bestimme, wann und wodurch die Bindung entsteht. Heutige Anerkennung darf nicht rückwirkend als

ursprüngliche Berechtigung vorausgesetzt werden; veränderte Regeln und relevante Fallunterschiede müssen berücksichtigt werden.

Reformulierungsaufgaben: Definiere „Ausstieg“, „legitimieren“, „Präzedenzfall“ und „Anspruch“ sowie die Vergleichbarkeitskriterien. Entscheide, ob soziale Geltung, institutionelle Berechtigung oder moralische Rechtfertigung untersucht wird. Benenne die zusätzliche Berechtigungsnorm und erläutere, warum bislang nicht begünstigte Mitglieder darunterfallen. Unterscheide Anhörung, Prüfung und Entscheidung zugunsten des Ausstiegs. Behandle Gnade, fehlerhafte Begünstigung und widerrufliche Ausnahmen als Grenzfälle. Begrenze schließlich die Aussage über die „wirksame Einheit“ auf den tatsächlich analysierten Überlieferungstyp.

2. H2 — Hermeneutisches Vorwurfs-Set. Der Einwand gegen eine Freiheitsentstehung durch Regelbruch belegt noch keinen besonderen Selbstbezug der Befragung. Adressierbarkeit, Handlungsfähigkeit, Selbstbewusstsein und Verantwortlichkeit sind verschiedene Zuschreibungen. Ein adressiertes Gebot beweist für sich weder reflektierte Freiheit noch exzentrische Positionalität.

H0/Gegendeutung: Die Befragung macht einen vorhandenen Selbstbezug sichtbar und verhandelt Verantwortlichkeit. Ihre erzählerische Funktion verlangt keinen neuen Modus der Selbstzuschreibung.

Zwei Alternativen: Erstens könnte gegenseitige Scham bereits den entscheidenden relationalen Selbstbezug markieren. Zweitens könnte die Erzählung einen Erkenntnis- und Statuswechsel gegenüber Gott entfalten, dessen Schwerpunkt auf veränderter normativer Stellung liegt.

Text- und Zeitprüfung: Genesis 2,16–17 enthält das Gebot; 3,2–3 gibt die erinnernde Antwort der Frau wieder. Diese Rede zeigt vor dem Essen eine sprachliche Bezugnahme auf eine Handlungsnorm. 3,7 stellt gegenseitige Scham vor die Befragung. In 3,9–13 stehen Handlungseingeständnis und Verantwortungsverschiebung nebeneinander. Zu klären ist, welchen zusätzlichen Selbstbezug die Befragung gegenüber Erinnerung und Scham erschließt. Vergleiche dabei getrennt, wem welche Fähigkeit zugeschrieben wird.

Prüfmethode und Quellenlage: Hermeneutisch sind Textdeckung, Erklärungsleistung und konkurrierende Lesarten abzuwägen; eine chronologische Folge beweist keine psychische Verursachung. Textuelle Gegenbelege müssen klar benannte Auslegungsansprüche treffen. Bloß fehlende Eindeutigkeit lässt die Lesart zunächst unbestätigt. Die einschlägigen Plessner-Passagen und vollständigen Vault-Vorgänger liegen nicht vor. Die Prüfung beruht auf den mitgelieferten Angaben.

Reformulierungsaufgaben: Grenze die These ausdrücklich gegenüber den Vorgängern vom 20./21.05.2026 ab. Benenne, welche Entstehungsbehauptung kritisch umgekehrt wird und welche Fähigkeit bereits vorausliegt. Eine durch Ansprache vermittelte Selbstbeziehung könnte sonst die ältere Erklärung von Selbstbewusstsein durch Fremdbeschreibung wiederholen. Setze frühere Freiheitsdeutungen nicht unbelegt mit Bewusstlosigkeit gleich. Bestimme den zusätzlichen Selbstbezug anhand der verglichenen Stellen und erkläre, welche Textmerkmale die spezifisch plessnersche Perspektive tragen. Kennzeichne diese als nachträgliches Interpretationsangebot. Unterscheide schließlich die Entstehung einer Fähigkeit von ihrer Ausübung und der ausdrücklich eingeforderten Rechenschaft.

3. H3 — Empirisches Vorwurfs-Set. Der Kernsatz behauptet sowohl einen Nutzen verzögerter Zuschreibung bei vorhandener Befundgeschichte als auch eine positive Interaktion. Ein bloßer Vorteil der Kombination genügt nicht: Zwei additive Haupteffekte könnten denselben Befund erzeugen.

H0/Gegendeutung: Getrennt zu prüfen sind das Ausbleiben eines praktisch relevanten Verzögerungsnutzens bei vorhandener Geschichte und das Ausbleiben einer positiven Interaktion. Beide betreffen unterschiedliche Vergleiche.

Zwei Alternativen: Erstens könnte die Revisionsgeschichte allein die Informationsgrundlage verbessern. Zweitens könnte verzögerte Zuschreibung allgemeine Konformität unabhängig von der Geschichte reduzieren. Beides könnte Ergebnisse verbessern, ohne den behaupteten besonderen Zusammenhang zu bestätigen.

Quellenlage: Nach den mitgelieferten verifizierten Hinweisen zeigt [Huber et al. \(2022\)](#) Statusbias; [Shah \(2023\)](#) untersucht bereits verzögerte Entblindung; [Liu et al. \(2024\)](#) findet keinen signifikanten Ankereffekt. Daraus folgt weder ein Beleg für diese Gruppenintervention noch ein Nachweis fehlender Ankerwirkung. Zeitversetzte Offenlegung allein trägt keinen Neuheitsanspruch; die Neuheit der spezifischen Kombination bleibt gesondert nachzuweisen. Eine Gesamtlektüre liegt dieser Bewertung nicht zugrunde.

Prüfmethode: Randomisiere auf Gruppenebene und berücksichtige die Abhängigkeit mehrerer Urteile derselben Gruppe. Gleiche Evidenzmenge, Aufgaben und Bearbeitungszeit an; variiere gezielt die Rekonstruierbarkeit der Revisionen. Verwende vorab validierte richtige und fehlerhafte Befunde sowie verblindete Ergebnisbewertung. Prüfe, ob die Revisionsspur Urheber verrät, und trenne Status von einschlägiger Expertise.

Zeitliche Logik: Fixiere das Primärurteil vor der Entblindung. Nachträgliche Wiedereröffnung und längerfristige Lernwirkungen benötigen eigene Messzeitpunkte. Auch erwartete spätere Zuschreibung kann bereits die erste Entscheidung beeinflussen.

Reformulierungsaufgaben: Operationalisiere richtige Korrekturen und irrtümliche Änderungen getrennt. Lege Interaktionsskala, Mindesteffekte, erforderliche Teststärke und die Bewertungsregel für zusätzliche Fehlkorrekturen vorab fest. Prüfe den behaupteten Statusmechanismus gesondert; bessere Entscheidungen allein identifizieren ihn nicht. Ein nichtsignifikantes Ergebnis bei breiten Unsicherheitsintervallen lässt die These unbestätigt. Hinreichend präzise Intervalle können praktisch relevante Vorteile ausschließen; eine belastbare Zunahme falscher Korrekturen wäre anhand der vorab festgelegten Regel als Gegenbefund zu beurteilen.

Reformulierung nach dem Kritischen Professor

H1 — Überlieferung als Einspruchsrecht

Kernsatz. Ein überlieferter Ausstiegsfall wird zu einem individuell beanspruchbaren Präzedenzgrund, wenn eine zuständige Instanz sich auf gleiche Prüfung vergleichbarer Fälle festlegt und Betroffenen erlaubt, diese Festlegung einzufordern. Seine Überlieferung bewahrt dann einen Prüfungsgrund; sie begründet für sich weder das Verfahrensrecht noch einen Anspruch auf dieselbe Entscheidung.

Begründung. Untersucht wird institutionelle Berechtigung zum Verlangen einer begründeten Prüfung, nicht die moralische Rechtfertigung eines Ausstiegs. Ausstieg bedeutet die Aufgabe eines bislang verbindlich verfolgten Ziels. Vier Momente bleiben getrennt: die frühere Entscheidung, ihre dokumentierte Weitergabe, die gegenwärtig geltende Selbstbindung und das heutige Verlangen. Als vergleichbar gelten Fälle nach vorab benannten Gründen, Zuständigkeit und Regelstand; geänderte Regeln verlangen eine eigene Begründung. Die normative Brücke liegt in ausdrücklich geltenden Verfahrensregeln und ihrem individuellen Anspruchscharakter. Die These erklärt, warum dieselbe Erzählung einmal bloße Orientierung und einmal eine einforderbare Prüfungsressource ist. Sie behauptet keine Ableitung moralischer Gleichheit aus Gewohnheit.

Änderung. Trennt Argument, individuelles Prüfungsrecht und Ausstiegsentscheidung; bezeichnet die zusätzliche Verfahrensnorm. Verzichtet auf einen allgemeinen Anspruch aus Tradition allein.

Prüfweg. Definitionsanalyse und vier Gegenmodelle: bloße Gewohnheit; objektive Bindung ohne Anspruch; allgemeines Prüfungsrecht ohne Tradition; ausdrückliche Selbstbindung mit individuellem

Verlangen. Gnade, rechtswidrige Begünstigung und widerrufliche Ausnahme werden getrennt. Die Abhängigkeit des Anspruchs von zusätzlicher Verfahrensgeltung ist offenzulegen; sie begrenzt die Originalität.

Scheiterpunkt. Die behauptete Unterscheidung scheitert an einem kohärenten Fall, in dem Betroffene nach den geltenden Regeln ausdrücklich gleiche Prüfung vergleichbarer Fälle verlangen dürfen, der überlieferte vergleichbare Fall dabei aber grundsätzlich kein relevanter Prüfungsgrund sein kann. Ein Modell objektiver Bindung ohne individuelles Recht widerlegt hingegen die stärkere Initialthese und ist in der Revision ausdrücklich zugelassen.

Gegendeutung. Eine Institution kann objektiv an frühere Entscheidungen gebunden sein, ohne den betroffenen Personen ein subjektives Recht auf Überprüfung einzuräumen. Dann besteht Präzedenzbindung, aber nicht der hier eng definierte individuell beanspruchbare Prüfungsgrund.

H2 — Das Selbst im Verhör

Kernsatz. Genesis 2–3 erzählt bereits vor dem Essen sprachliche Bezugnahme auf Gebot und Handlung; die spätere Befragung fordert erstmals in dieser Erzählsequenz eine ausdrückliche Rechenschaft über die vollzogene verbotene Handlung. Als plessnersch inspirierte Lesart lässt sich dieser Übergang als öffentlich eingeforderte Selbstzuschreibung verstehen, ohne darin die Entstehung von Handlungsfähigkeit oder Selbstbewusstsein zu behaupten.

Begründung. Vier Textstationen sind am Primärtext der Einheitsübersetzung geprüft: Gen 2,16–17 adressiert das Gebot an den Menschen; Gen 3,2–3 zeigt die erinnernde Normrede der Frau; Gen 3,7–8 erzählt Nacktheitserkenntnis und Verbergen vor der Befragung; Gen 3,9–13 verbindet Antworten über das eigene Handeln mit der Verlagerung von Verantwortung auf andere. Der zusätzliche Aspekt ist die gegenüber einem Fragenden eingeforderte Rechenschaft, nicht ein erstmals existierendes Ich. Der Begriff öffentlich bedeutet hier gegenüber einem anderen artikuliert, nicht Öffentlichkeit im modernen politischen Sinn. Die Perspektive Plessners ist eine nachträgliche Heuristik; einschlägige Originalpassagen wurden noch nicht geprüft.

Änderung. Rechenschaft statt Ich-Entstehung; vier geprüfte Textstationen, konkurrierende Deutungen und ungesicherter Plessner-Bezug ausdrücklich markiert.

Prüfweg. Vergleiche Fähigkeiten und jeweilige Figuren getrennt an den vier Textstationen. Prüfe die Konkurrenzlesarten Scham, Erkenntnis und normative Stellung auf gleiche Textabdeckung. Kein Schluss von Erzählfolge auf psychische Kausalität oder historische Anthropogenese.

Scheiterpunkt. Die enge Auslegung verliert ihren behaupteten Zusatznutzen, wenn eine konkurrierende Lesart den Wechsel von Normrede und Verbergen zur ausdrücklich eingeforderten Rechenschaft mit weniger Zusatzannahmen ebenso genau erklärt. Ein vor der Befragung belegtes Selbstverhältnis ist kein Gegenbeweis: Es ist Bestandteil der revidierten Lesart.

Abgrenzung. Die Reservoir-Texte vom 20./21.05.2026 erklären Selbstentstehung durch Schuld oder Fremdbeschreibung. Die Revision kehrt diese Entstehungsbehauptung kritisch um: Sie setzt vorhandene Fähigkeiten voraus und untersucht ihren situativen Gebrauch. Die Genesis-Notiz ist historisch kein Virgin Node.

Gegendeutung. Die Befragung kann allein eine schon bestehende Verantwortlichkeit sichtbar machen; Scham und Nacktheitserkenntnis oder die geänderte Stellung gegenüber Gott können den erzählerischen Schwerpunkt bilden.

H3 — Die namenlose Fehlerspur

Kernsatz. Bei der gemeinsamen Prüfung vorab als richtig oder fehlerhaft validierter Forschungsbefunde verbessert eine bis zum fixierten Ersturteil verzögerte persönliche Zuschreibung die Rate richtiger Korrekturen stärker, wenn die sachliche Revisionsgeschichte rekonstruierbar bleibt. Behauptet werden ein positiver Verzögerungsnutzen bei vorhandener Geschichte und eine positive Interaktion beider Bedingungen, ohne eine vorab als relevant bestimmte Zunahme falscher Korrekturen.

Begründung. Eine überprüfbare Geschichte früherer Behauptungen, Belege und Änderungen kann Gründe zugänglich halten, während zeitweilig verdeckte Namen vorschnelle persönliche Bindungen verringern könnten. Das ist eine unbestätigte Vermutung über eine gemeinsame Entscheidungsaufgabe, keine generelle Empfehlung für anonyme Wissenschaft. Verzögerte Offenlegung wurde bereits von Shah 2023 untersucht; weder Anonymität noch das zeitliche Verfahren beanspruchen Neuheit. Der zu prüfende Rest betrifft ihre Interaktion mit der Rekonstruierbarkeit einer sachlichen Befundgeschichte. Ein Statusmechanismus folgt nicht schon aus einem Leistungsvorteil.

Änderung. Getrennte Endpunkte und vorab fixiertes Urteil; positiver einfacher Effekt und echte Interaktion auf festgelegter Skala; explizite Gegenbefunde und noch offene Schwellenkalibrierung.

Prüfweg. Ein Pilot prüft Durchführbarkeit, Durchsickern persönlicher Identität und Verständlichkeit, nicht die Bestätigung der These. Vor dem Hauptversuch werden minimale relevante Vorteile, Toleranzgrenze falscher Änderungen, Fallzahl und Analyse festgelegt; die Grenzwerte sind noch nicht empirisch kalibriert. Gruppierung und mehrere Aufgaben pro Gruppe gehen in die Unsicherheitsanalyse ein. Änderungen nach Entblindung und späteres Lernen sind separate Folgevariablen.

Scheiterpunkt. Für die vorab festgelegte Population und Aufgabengruppe trifft ein hinreichend präzises Ergebnis die starke Kombinationshypothese, wenn der minimale Verzögerungsnutzen bei vorhandener Geschichte oder die minimale positive Interaktion ausgeschlossen wird. Eine Zunahme falscher Korrekturen über die festgelegte Toleranzgrenze verletzt die Sicherheitsbedingung. Breite Intervalle oder gescheiterte Verblindung liefern keine Bestätigung und erlauben nicht ohne Weiteres eine Widerlegung der zugrunde liegenden Vermutung.

Operationalisierung. Faktorielles 2x2-Experiment mit Randomisierung vollständiger Gruppen: sachliche Revisionsgeschichte rekonstruierbar versus dieselben Evidenzstücke ohne verknüpfte Änderungsschritte; persönliche Zuschreibung vor versus erst nach fixierter gemeinsamer Erstentscheidung. Evidenzmenge, Aufgaben, Zeit, Expertiseinformation und erwartete spätere Verantwortlichkeit werden angeglichen. Persönliche Statusmerkmale sind von aufgabenrelevanter Expertise zu trennen. Verblindete unabhängige Prüfer validieren die richtige Lösung, bevor Gruppen teilnehmen. Für fehlerhafte Ausgangsbefunde wird die Rate richtiger Korrekturen gemessen; für richtige Ausgangsbefunde separat die Rate falscher Änderungen. Die primäre Skala sind absolute Risikodifferenzen. Die Interaktion ist der Verzögerungseffekt mit rekonstruierbarer Geschichte minus derselbe Effekt ohne rekonstruierbare Geschichte. Keine Auswertung allein der Änderungszahl.

Gegendeutung. Früh sichtbare persönliche Verantwortlichkeit kann Sorgfalt fördern; Expertise kann an Personen gebunden und sachlich nützlich sein. Eine Revisionsgeschichte kann die Erkennung ihrer Autoren erleichtern oder nachträgliche Rechtfertigungen begünstigen. Dann verbessert verzögerte Zuschreibung die Qualität nicht oder verschlechtert sie.

Unabhängige Bewertung und Ranking

Die Bewertung entstand in einem frischen GPT-6-Astra/max-Kontext ohne Vorgabe einer bevorzugten Hypothese. Kriteriennamen und Anker folgen der Projektkonfiguration; die tatsächlichen Gewichte ergeben die Normierung 0,9.

Kriterium	H1	H2	H3
Originalität	4	5	6
Falsifizierbarkeit	6	6	8
Begriffliche Klarheit	8	8	8
Tiefe	6	7	7
Forschungsrelevanz	6	6	8
Interdisziplinäre Anschlussfähigkeit	6	6	8
Vault-Anschluss	5	7	5
Antinomie-Test	5	6	7
Publikationsmöglichkeit	6	6	7
Summe /90	51	56	63

H1 — Überlieferung als Einspruchsrecht

Themenklasse: ethisch_praktische_hypothese. **Methodentyp:** begriffsanalytisch. **Reichweite:** these. Der Kandidat formuliert eine abgegrenzte Aussage über das Verhältnis von Überlieferung und individuellem Prüfungsrecht mit einem eigenen Gegenmodelltest. Er erweitert die vorhandene Ausstiegs- und Provenienzzlinie um eine institutionelle Anspruchsbedingung.

Pflichttest: Dominant ist die Begriffsanalyse institutioneller Anspruchsbedingungen; es gelten die Obergrenzen Falsifizierbarkeit 7, Forschungsrelevanz 9 und Interdisziplinäre Anschlussfähigkeit 8, sonst 10. Der Definitions-Schärfe-Check ist bestanden, weil Entscheidung, Überlieferung, Selbstbindung und individuelles Verlangen getrennt sowie Vergleichbarkeit und Geltungsbedingungen ausdrücklich bestimmt werden.

- **Originalität (4/10):** Die ausdrücklich vorausgesetzte Verfahrensnorm übernimmt den entscheidenden Begründungsschritt, sodass der eigenständige Beitrag hauptsächlich in der Trennung von Überlieferung und Anspruchsgrundlage liegt. Die externe Neuheit dieser begrenzten Unterscheidung ist ungeprüft.
- **Falsifizierbarkeit (6/10):** Die vier Gegenmodelle und der mögliche Fall eines Gleichprüfungsrechts ohne Relevanz früherer Entscheidungen ergeben einen konkreten begrifflichen Prüfpfad. Die Relevanz eines Präzedenzfalls muss allerdings unabhängig von der vorausgesetzten Gleichprüfungsregel bestimmbar bleiben.
- **Begriffliche Klarheit (8/10):** Prüfungsgrund, individuelles Verfahrensrecht und Anspruch auf dieselbe Entscheidung werden ausdrücklich auseinandergehalten. Noch schärfer abzugrenzen ist der eigenständige Begründungsbeitrag eines Präzedenzgrundes gegenüber einem bloßen Hinweis auf inkonsistente Regelanwendung.
- **Tiefe (6/10):** Die Unterscheidung zwischen tradierten Gründen und gegenwärtiger Berechtigung berührt die institutionelle Entstehung individueller Ansprüche. Die tragende Verfahrensgeltung wird jedoch vorausgesetzt und selbst nicht erklärt.

- **Forschungsrelevanz (6/10):** Subjektive Verfahrensrechte, institutionelle Selbstbindung und Präzedenzrelevanz bieten konkrete rechtsphilosophische Anschlussfragen. Eine aktuelle Fachkontroverse, zu der die Unterscheidung nachweislich etwas beiträgt, ist im Material nicht dokumentiert.
- **Interdisziplinäre Anschlussfähigkeit (6/10):** Rechtsphilosophie, Sozialontologie und Institutionentheorie können unmittelbar an individuelle Ansprüche, soziale Akte und institutionelle Selbstbindung anschließen.
- **Vault-Anschluss (5/10):** Der Ausstiegs- und Provenienzvorgänger sowie Reinachs soziale Akte liefern benannte Anschlussstellen. Die vorliegenden Angaben enthalten jedoch keine ausgearbeitete Rekonstruktion dieser Verbindungen oder des zusätzlichen Luhmann-Bezugs.
- **Antinomie-Test (5/10):** Die Gegenthese, gleiche Prüfung könne ausschließlich auf gegenwärtigen Regeln beruhen und frühere Entscheidungen nur illustrieren lassen, trifft den behaupteten Beitrag der Überlieferung. Die im Kandidaten genannte objektive Bindung ohne Individualanspruch ist dagegen bereits zugelassen und erzeugt keinen verbleibenden Widerspruch.
- **Publikationsmöglichkeit (6/10):** Die abgegrenzte Argumentation bietet eine Grundlage für einen begriffsanalytischen Fachaufsatz. Dafür müssen die einschlägige Präzedenz- und Anspruchstheorie sowie der zusätzliche Erkenntnisgewinn gegenüber der vorausgesetzten Verfahrensnorm ausgearbeitet werden.

Stärkste Gegenthese: Auch ein individuell einforderbares Recht auf gleiche Prüfung kann ausschließlich an geltende Regeln und gegenwärtige Sachgründe gebunden sein; überlieferte Entscheidungen liefern dann Hinweise auf mögliche Inkonsistenzen, aber keinen eigenständigen Präzedenzgrund.

H2 — Das Selbst im Verhör

Themenklasse: anthropologische_hypothese. **Methodentyp:** hermeneutisch. **Reichweite:** these. Die vier Textstationen tragen eine abgegrenzte Auslegung, deren zusätzlicher Erklärungswert durch konkurrierende Deutungen geprüft werden kann. Der ungeprüfte Plessner-Bezug begrenzt den philosophischen Ausbau dieser eigenständig prüfbaren Textthese.

Pflichttest: Für die dominante hermeneutische Auslegung situativer Rechenschaft gelten die Obergrenzen Falsifizierbarkeit 6 und Antinomie-Test 7, sonst 10. Der Quellen-Adäquatheits-Check ist mit den vier dokumentiert geprüften Stationen Gen 2,16–17; 3,2–3; 3,7–8 und 3,9–13 bestanden; die ergänzende Plessner-Heuristik bleibt ohne Originaltextprüfung.

- **Originalität (5/10):** Die Verschiebung von Selbstentstehung zu situativ eingeforderter Rechenschaft setzt einen unterscheidbaren Akzent gegenüber den beiden Reservoir-Vorgängern. Ihre Neuheit innerhalb der Genesis-Exegese und Plessner-Forschung ist ungeprüft.
- **Falsifizierbarkeit (6/10):** Die behauptete Erzählsequenz lässt sich an vier dokumentiert geprüften Textstationen kontrollieren. Der zusätzliche Auslegungsnutzen kann daran scheitern, dass eine konkurrierende Deutung dieselben Übergänge mit weniger Zusatzannahmen erklärt.
- **Begriffliche Klarheit (8/10):** Vorhandene Handlungsfähigkeit, vorausgehendes Selbstverhältnis und später eingeforderte Rechenschaft werden klar getrennt; öffentlich wird ausdrücklich als gegenüber einem anderen artikuliert bestimmt. Der spezifisch plessnersche Gehalt dieser Unterscheidung bleibt unterbestimmt.
- **Tiefe (7/10):** Die Unterscheidung zwischen vorhandenen Fähigkeiten und ihrer durch andere eingeforderten Ausübung berührt das Verhältnis von Selbstbezug und Verantwortungszuschreibung. Eine weitergehende Erklärung dieses Verhältnisses wird durch die Textsequenz allein noch nicht geleistet.

- **Forschungsrelevanz (6/10):** Rechenschaft, Selbstzuschreibung und Genesis-Auslegung ergeben einen konkreten Gegenstand für philosophische Anthropologie und Exegese. Ein Beitrag zu einer aktuellen Fachkontroverse oder zu Plessners tatsächlicher Argumentation ist bisher nicht dokumentiert.
- **Interdisziplinäre Anschlussfähigkeit (6/10):** Biblische Exegese, philosophische Anthropologie und Erzähltheorie können unmittelbar an Figurenrede, Verantwortungszuschreibung und die Abfolge der Textstationen anschließen.
- **Vault-Anschluss (7/10):** Die Revision verbindet die Genesis-Notiz mit einer ausdrücklich beschriebenen Korrektur der beiden Reservoir-Thesen zu Alterität und Schuld. Dieser konkrete argumentative Anschluss belegt jedoch keine drei dicht ausgearbeiteten Theorieknoten.
- **Antinomie-Test (6/10):** Eine Schuld- und Gerichtslesart kann die Antworten als Rechtfertigung und Verantwortungsabwehr erklären, ohne öffentliche Selbstzuschreibung als zusätzlichen Erklärungsbegriff einzuführen. Diese Konkurrenz ist textnah, bleibt aber teilweise mit der revidierten These vereinbar.
- **Publikationsmöglichkeit (6/10):** Ein quellengebundener Aufsatz über die Rechenschaftsszene besitzt einen klar begrenzten Gegenstand. Für einen eigenständigen philosophischen Beitrag fehlen die Plessner-Originalstellen und die Einordnung in die exegetische Forschung.

Stärkste Gegenthese: Die Befragung inszeniert die bereits durch Übertretung, Scham und verletzte Gottesbeziehung bestimmte Verantwortlichkeit; eine Schuld- und Gerichtslesart erklärt die Antworten einschließlich ihrer Verantwortungsabwehr ebenso genau, sodass öffentlich eingeforderte Selbstzuschreibung keinen zusätzlichen Erklärungsgewinn liefert.

H3 — Die namenlose Fehlerspur

Themenklasse: epistemologisch_systemtheoretisch_hypothese. **Methodentyp:** empirisch_pruefbar.

Reichweite: these. Der Kandidat behauptet einen abgegrenzten Interaktionseffekt samt Mindestnutzen und Fehlkorrekturgrenze und benennt einen eigenen experimentellen Widerlegungsweg. Pilot und Folgevariablen ergänzen dieses Untersuchungsdesign, ergeben aber noch keine Architektur mehrerer eigenständiger Teilthesen.

Pflichttest: Die kausale Interaktionsbehauptung ist dominant empirisch prüfbar; es gelten die Obergrenzen Antinomie-Test 7 und Tiefe 9, sonst 10. Der Operationalisierungs-Check ist bestanden, weil beide Faktoren, richtige und falsche Änderungen sowie die Interaktion auf absoluter Risikodifferenzskala messbar definiert sind; die offene Schwellenkalibrierung begrenzt die Ausführungsreife.

- **Originalität (6/10):** Als vorläufig eigenständiger Beitrag verbleibt die positive Interaktion zwischen Zuschreibungsverzögerung und rekonstruierbarer Befundgeschichte. Verzögerte Namensoffenlegung ist durch die dokumentierte Prüfung von Shah 2023 bereits bekannt, und die begrenzte Recherche sichert keine Priorität der Interaktionsbehauptung.
- **Falsifizierbarkeit (8/10):** Das gruppenrandomisierte 2×2-Design trennt den Verzögerungsnutzen bei vorhandener Geschichte, die Interaktion und die Zunahme falscher Korrekturen mit ausdrücklichen Scheiterbedingungen. Konkrete Population, Aufgaben und kalibrierte Mindest- und Toleranzwerte fehlen im vorliegenden Entwurf für einen vollständig festgelegten Test.
- **Begriffliche Klarheit (8/10):** Einfacher Effekt, Interaktion, richtige Korrekturen und falsche Änderungen sind auf einer ausdrücklich benannten Auswertungsskala getrennt. Rekonstruierbarkeit und wirksame Identitätsverdeckung müssen noch durch konkrete Materialien und Manipulationsprüfungen abgesichert werden.
- **Tiefe (7/10):** Die Trennung persönlicher Zuschreibung von sachlicher Nachvollziehbarkeit berührt Bedingungen kollektiver Erkenntnis Korrektur. Der Versuch prüft zunächst einen begrenzten

Leistungseffekt und identifiziert den vermuteten Statusmechanismus noch nicht.

- **Forschungsrelevanz (8/10):** Die dokumentierten Arbeiten von Huber, Shah und Pleskac begründen einen konkreten Anschluss an Autorstatus und Verblindung im Peer Review. Die getrennte Messung sachlich richtiger und falscher Revisionen richtet diesen Anschluss auf eine Ergebnisgröße, die aus Status- oder Fairnessunterschieden allein nicht folgt.
- **Interdisziplinäre Anschlussfähigkeit (8/10):** Wissenschaftssoziologie, experimentelle Sozialpsychologie, Metawissenschaft und Informationswissenschaft erhalten mit Statusinformation, Gruppenentscheidungen, Korrekturgüte und rekonstruierbaren Versionsverläufen jeweils konkrete Ansatzpunkte.
- **Vault-Anschluss (5/10):** Vertrauen und die vorhandene Provenienzlinie bieten zwei erkennbare Anschlussstellen. Die vorgelegten kurzen Anker enthalten jedoch keine ausgearbeitete Herleitung des experimentellen Interaktionseffekts aus diesen Beständen.
- **Antinomie-Test (7/10):** Die Gegenthese, frühe persönliche Zuschreibung erhöhe kooperative Sorgfalt und mache die Verzögerung wirkungslos oder schädlich, betrifft denselben messbaren Endpunkt. Die dokumentierte Gegenperspektive von Leek 2011 motiviert diese Möglichkeit, ohne sie für die geplante Gruppenaufgabe bereits zu belegen.
- **Publikationsmöglichkeit (7/10):** Die vorgelegten Peer-Review-Studien in PLOS ONE zeigen einen plausiblen Publikationsanschluss für das Experiment. Eine ausgereifte empirische Einreichung setzt Pilotierung, festgelegte Entscheidungsgrenzen und Ergebnisse voraus; eine Annahmewahrscheinlichkeit ist nicht belegt.

Stärkste Gegenthese: Bei gleicher Expertiseinformation und gleicher erwarteter späterer Verantwortlichkeit kann früh sichtbare persönliche Zuschreibung unmittelbare kooperative Sorgfalt fördern; eine rekonstruierbare Revisionsgeschichte kann diesen Vorteil verstärken, sodass die Verzögerung keinen positiven Nutzen oder keine positive Interaktion erzeugt.

Auswahl und Qualitätsentscheid

H3: Die namenlose Fehlerspur (63/90) → H2: Das Selbst im Verhör (56/90) → H1: Überlieferung als Einspruchsrecht (51/90).

H3 gewinnt aufgrund des gruppenrandomisierten Interaktionsdesigns mit getrennten Endpunkten für richtige Korrekturen und falsche Änderungen und wird für die Expertenrunde gewählt; H2 ist die Nebenlinie, H1 belegt Platz 3. Rechnerisch geprüft ergeben sich bei Gewichtssumme 10,0 und kaufmännischer Rundung nach Multiplikation mit 0,9 die Werte H3: 70,4 → 63,36 → 63, H2: 62,6 → 56,34 → 56 und H1: 56,6 → 50,94 → 51, womit die Rangfolge bereits vor Rundung eindeutig ist. H3s wichtigste Grenze sind die noch fehlenden konkreten Testfestlegungen und die ungeprüfte Übertragbarkeit auf reale Forschungspraxis. Der Neuheitsstatus bleibt bei allen Kandidaten vorläufig; bei H3 betrifft er ausschließlich die Interaktionsbehauptung und nicht die bereits dokumentierte verzögerte Offenlegung. 9er- oder 10er-Scores sind auf der vorliegenden Grundlage nicht ausreichend belegt, insbesondere angesichts offener Testparameter sowie fehlender belastbarer Neuheitsnachweise, dichter Theorieanschlüsse und gesicherter Publikationsperspektiven.

Hauptlinie: Die namenlose Fehlerspur. **Nebenlinie:** Das Selbst im Verhör. Nur die Hauptlinie wird in Runde 2 fortgesetzt. **Panel:** Kant, Popper, Wittgenstein, Plessner, Cassirer, Friston, Luhmann; Konfiguration epistemologisch_systemtheoretisch_hypothese. Die Rotationssperre von drei Tagen ist beim Abstand zum letzten vollständigen Lauf vom 05.10. eingehalten.

Expertenrunde 1 — Hauptlinie

Hypothese. Die namenlose Fehlerspur (63/90).

Kernsatz. Bei der gemeinsamen Prüfung vorab als richtig oder fehlerhaft validierter Forschungsbefunde verbessert eine bis zum fixierten Ersturteil verzögerte persönliche Zuschreibung die Rate richtiger Korrekturen stärker, wenn die sachliche Revisionsgeschichte rekonstruierbar bleibt. Behauptet werden ein positiver Verzögerungsnutzen bei vorhandener Geschichte und eine positive Interaktion beider Bedingungen, ohne eine vorab als relevant bestimmte Zunahme falscher Korrekturen.

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

Separat gestartete Expertenkontexte ohne andere Gutachten. KI-Rekonstruktionen; keine Originaläußerungen der genannten Personen.

Gutachten 1 — Kant

Diese KI-Rekonstruktion der Kant-Perspektive unterscheidet zunächst die Übereinstimmung eines Urteils mit der vorgegebenen Lösung von seiner Berechtigung durch zugängliche Gründe. Das Experiment kann jene Übereinstimmung messen; ob beide Versuchsbedingungen dieselbe sachliche Urteilsaufgabe stellen, bedarf jedoch einer eigenen Prüfung. Eine Revisionsgeschichte verbindet nicht bloß Darstellungen, sondern kann anzeigen, welche Behauptung zurückgenommen wurde, welcher Beleg weiterhin gilt und welche Schlussfolgerung von einem verworfenen Schritt abhängt. Gleiche Evidenzstücke gewährleisten deshalb noch keine Gleichheit der verfügbaren Rechtfertigungsgründe. Der Einwand lautet: Wenn erst die verknüpfte Geschichte erkennen lässt, welcher Befund maßgeblich ist, kann ein nach dem Prüfermaßstab falsches Gruppenurteil unter den tatsächlich zugänglichen Informationen durchaus vernünftig sein. Damit würde der vorvalidierte Lösungsmaßstab auch dort uneingeschränkt angewandt, wo seine sachlichen Voraussetzungen den Urteilenden fehlen. Das widerlegt die eng gefasste Interaktionshypothese nicht, begrenzt aber ihre Auslegung als Verbesserung von Fehlerkorrektur bei gleicher Evidenz. Zur Schärfung sollte die Vorvalidierung für jede Darstellungsbedingung gesondert dokumentieren, ob die richtige Lösung aus den dort verfügbaren Informationen begründbar ist; der Status einzelner Evidenzstücke als aktuell oder zurückgenommen müsste in beiden Bedingungen erkennbar bleiben, während nur ihre Verknüpfung zu Änderungsschritten variiert. Ergänzend könnten die Gruppen mit ihrem Ersturteil den tragenden Schluss und einen möglichen Widerlegungsgrund festhalten. So würde prüfbar, ob die Geschichte bloß zum richtigen Ergebnis führt oder auch dessen sachliche Begründung erschließt. Schließlich bleibt der formale Anspruch, Gründe unabhängig vom Ansehen einer Person beurteilen zu können, vom empirischen Befund zu trennen, welche Einrichtung seine Ausübung erleichtert.

Gutachten 2 — Popper

Diese KI-Rekonstruktion der Popper-Perspektive setzt beim Wahrheitsmaßstab an: Die Unabhängigkeit der Prüfer schützt vor bestimmten Verzerrungen, macht ihr Urteil aber nicht unfehlbar. Wer die sachliche Prüfung von persönlicher Autorität lösen will, muss diese Forderung auch gegenüber jenen Personen erheben, die richtige und fehlerhafte Ausgangsbefunde festlegen. Mein konkreter Einwand lautet: Falls diese Einteilung lediglich auf Expertenkonsens beruht, könnte eine sachlich berechnete Abweichung als falsche Korrektur zählen. Eine positive Interaktion würde dann möglicherweise größere Übereinstimmung mit dem Bewertungsschlüssel anzeigen, ohne bessere Fehlerkorrektur zu belegen. Der Entwurf erläutert die unabhängige Validierung, aber noch nicht, wodurch ihr Maßstab seinerseits kritisierbar bleibt. Methodisch sauber wäre zunächst ein Aufgabenbestand, dessen Fehlerstatus durch nachrechenbare Ergebnisse, explizite Widersprüche oder unabhängig reproduzierbare Befunde begründet wird. Für unerwartete Einwände gegen diesen Status braucht es ein vorab festgelegtes Nachprüfungsverfahren mit dokumentierten Entscheidungskriterien, dessen Prüfer die Versuchsbedingung nicht kennen; begründete Neuklassifikationen müssen für sämtliche Versuchsarme gleichermaßen gelten. Die ursprüngliche Auswertung und eine Auswertung nach begründeter Neuklassifikation sollten getrennt berichtet werden. Der entscheidende Schwellentest vergleicht dann Verzögerungsnutzen und Interaktion mit den festgelegten Mindestwerten; die Fehlerkorrekturgrenze ist zusätzlich einzuhalten. Liegt die obere Konfidenzgrenze unter einem dieser Mindestwerte, scheitert die starke Kombination am vorgesehenen Test; unentschiedene Ergebnisse bestätigen sie nicht. Der Preis dieser Schärfung ist eine zunächst engere Reichweite: Lösbare Prüfaufgaben erfassen nicht ohne Weiteres die offene Unsicherheit wirklicher Forschung.

Gutachten 3 — Wittgenstein

KI-Rekonstruktion der vorgegebenen Perspektive: Eine zurückgenommene Behauptung wird anders gebraucht als eine geltende. Das ist für diesen Versuch keine Nebensache. Mit einer rekonstruierbaren Geschichte können die Teilnehmer erkennen, was jemand behauptete, was später verworfen wurde und worauf sich die gegenwärtige Prüfung richtet. Dieselben Sätze ohne ihre Verknüpfungen müssen deshalb nicht dieselbe sachliche Ausgangslage herstellen. Stellen Sie sich vor: Ein Messwert wird berichtet, der alte bleibt dokumentiert. Wer beide Zahlen nebeneinander erhält, muss womöglich erst entscheiden, welche überhaupt zur Prüfung steht. Wer den Verlauf erhält, kann bereits über ihre Berechtigung urteilen. Wenn erst die Geschichte die Prüfaufgabe klärt, vergleichen Sie teilweise verschiedene Tätigkeiten unter dem gemeinsamen Namen Fehlerkorrektur. Mein Einwand betrifft deshalb die Rede von gleicher Evidenz: Die Anzahl der Stücke sichert keine Gleichheit ihres Gebrauchs. Das widerlegt den möglichen Interaktionseffekt nicht. Es beschränkt, was dieser Effekt über gemeinsam zugängliche Gründe aussagt. Die konkrete Schärfung: Kennzeichnen Sie in allen vier Bedingungen dieselbe zu prüfende Behauptung und welche Angaben als aktuell vertreten oder als zurückgezogen vorgelegt werden. Variieren Sie anschließend, ob die Übergänge zwischen früheren Fassungen nachvollziehbar sind. Prüfen Sie im Pilot durch kurze Wiedergaben, ob die Teilnehmer dieselbe Behauptung beurteilen. Halten Sie außerdem eine knappe Begründung des Gruppenurteils fest. So lässt sich unterscheiden, ob eine Gruppe einen sachlichen Fehler berichtet oder zunächst ein missverstandenes Prüfmaterial ordnet. Die vorab validierte Lösung bleibt ein brauchbarer Maßstab. Sie beantwortet aber nicht schon die Frage, was die Teilnehmer beim Korrigieren tun.

Gutachten 4 — Plessner

Diese KI-Rekonstruktion nach dem vorgegebenen Plessner-Profil setzt bei einer eigentümlichen Vermittlung an: Auch eine namenlose Revisionsgeschichte kann jemanden erscheinen lassen, der Behauptungen aufstellt, sich berichtigt oder an ihnen festhält. Die persönliche Bindung an einen Befund kann bei verdecktem Namen schwächer werden und sich zugleich am dargestellten Umgang mit Irrtümern neu bilden. Darin liegt ein anthropologisch interessanter Zug des Entwurfs: Distanz zum Urteil entsteht innerhalb einer gestalteten Beziehung zu anderen, deren Identität vorläufig offenbleibt. Mein Einwand betrifft die Trennung zwischen sachlicher Geschichte und persönlicher Zuschreibung. Eine geordnete Folge von Änderungen kann Verlässlichkeit, Einsicht oder Hartnäckigkeit nahelegen, obwohl Expertiseangaben gleichgehalten werden und keinerlei Name durchsickert. Der Geschichtsfaktor verändert dann möglicherweise auch die zugeschriebene Gestalt eines Gegenübers. Der faktorielle Versuch kann weiterhin den vorgesehenen Verfahrenseffekt bestimmen; dessen anthropologische Interpretation muss diese zusätzliche Vermittlung berücksichtigen. Zur Schärfung sollte ein eigener Materialpilot prüfen, ob und wem die Teilnehmenden die dokumentierten Revisionen zurechnen und welche Verlässlichkeit sie daraus ableiten. Insbesondere muss festgelegt werden, ob frühere Korrekturen von denselben Urhebern oder von fremden Prüfern stammen; diese Information ist zwischen den Zuschreibungsbedingungen konstant zu halten. Eine getrennte Auswertung der Gruppengespräche könnte ergänzend untersuchen, wann Gründe übernommen und wann vermutete Personen beurteilt werden. Ich würde die sachliche Nachvollziehbarkeit daher als institutionelle Bedingung möglicher Distanzierung fassen. Die gemessene Korrekturleistung gibt über diese Stellung zum fremden und eigenen Urteil nur begrenzt Auskunft.

Gutachten 5 — Cassirer

Die namenlose Fehlerspur erhält in dieser ausdrücklich als KI-Rekonstruktion angelegten Cassirer-Perspektive ihre Bedeutung durch einen Wechsel der symbolischen Träger wissenschaftlicher Autorität: Der Personenname tritt zeitweilig zurück, während die geordnete Folge von Behauptung, Einwand und Revision selbst Geltung vermitteln kann. Eine Revisionsgeschichte bewahrt deshalb nicht einfach dieselben Inhalte; sie stellt Beziehungen her, durch die überhaupt erst erkennbar wird, was als Verbesserung, Widerlegung oder bloßer Wechsel zählt. Darin liegt die philosophische Stärke des Entwurfs: Wissenschaftliche Objektivität erscheint als Leistung nachvollziehbarer Verknüpfungen, deren Gelingen nicht an eine bestimmte Person gebunden ist. Mein konkreter Einwand betrifft jedoch die noch offene Gestalt dieser Verknüpfungen. Chronologische Rekonstruierbarkeit, argumentative Abhängigkeit und die Erzählung eines fortschreitenden Erkenntnisgewinns sind verschiedene Leistungen; das vorgesehene Material könnte sie gemeinsam verändern. Dann wäre die Interaktion experimentell bestimmbar, ihre Bedeutung aber unklar: Die Gruppen könnten Gründen folgen oder der stillen Autorität einer als Fortschritt dargestellten Geschichte. Zur Schärfung sollten die Materialien Zeitfolge und Begründungsrelationen ausdrücklich unterscheiden, auch verworfene, wiederaufgenommene und weiterhin strittige Schritte enthalten und keine spätere Version allein durch ihre Darstellung als überlegen auszeichnen. Eine ergänzende Prüfung könnte bei gleicher Rekonstruierbarkeit eine chronologische Darstellung mit einer als Fortschritt gerahmten Darstellung vergleichen. Zusätzlich ließe sich nach dem fixierten Ersturteil erfassen, welche Begründungsrelationen die Gruppen tatsächlich rekonstruiert haben. Das erweitert den Leistungstest um die Frage, welche symbolische Ordnung den Zugang zu Gründen trägt, ohne einen möglichen Vorteil verdeckter Namen schon als Befreiung von Autorität zu deuten.

Gutachten 6 — Friston

Diese KI-Rekonstruktion der Friston-Perspektive versteht die Kombination als Eingriff in zwei Komponenten eines generativen Modells: Persönliche Zuschreibung könnte die Präzision sozial vermittelter Erwartungen verändern, rekonstruierbare Geschichte dagegen die Einschätzung, welche Evidenzstücke voneinander abhängig sind. Dieselbe Evidenzmenge bedeutet deshalb keineswegs denselben Informationsgehalt; mehrere Revisionen derselben Beobachtung dürfen nicht wie unabhängige Bestätigungen verrechnet werden. Das free energy principle allein legt weder diese Modellstruktur noch das Vorzeichen des Leistungseffekts fest. Mein konkreter Einwand betrifft die theoretische Bedeutung der positiven Interaktion: Auf der Risikodifferenzskala kann sie bereits durch eine nichtlineare Umsetzung unabhängiger Einflussgrößen in Entscheidungen entstehen. Schon ein Modell mit additiven Effekten auf die Log-Odds einer richtigen Korrektur kann eine solche Interaktion erzeugen, ohne eine spezifische Kopplung beider Eingriffe im Inferenzprozess. Der vorab definierte Leistungstest bleibt damit gültig; sein Bestehen würde diese stärkere Erklärung jedoch nicht tragen. Zur Schärfung sollte das Experiment ein ergänzendes Prozessmodell vorab spezifizieren, das Evidenzabhängigkeit, Präzisionsgewichtung und Entscheidungsschwelle getrennt repräsentiert. Aufgaben mit bekannten Abhängigkeiten zwischen Belegen sowie individuelle Wahrscheinlichkeitsurteile vor der Gruppendiskussion und eine gemeinsame Wahrscheinlichkeitseinschätzung zur Befundrichtigkeit vor der Entscheidung würden diese Unterscheidung unterstützen. Zu vergleichen wären ein Modell unabhängiger Beiträge und ein Modell, in dem Zuschreibung die Nutzung der Geschichte verändert; simulierte Daten sollten zunächst zeigen, ob die Parameter unterscheidbar sind. Die Risikodifferenzen bleiben die primären Endpunkte. Unveränderte Wahrscheinlichkeitseinschätzungen bei veränderten Korrekturentscheidungen sprächen eher für eine verschobene Entscheidungsschwelle.

Gutachten 7 — Luhmann

KI-Rekonstruktion: Die Versuchsanordnung entkoppelt zeitweise persönliche Namen von der Anschlussfähigkeit früherer Mitteilungen. Die angebotene Revisionsgeschichte kann zum Gedächtnis der Gruppenkommunikation werden, soweit diese auf frühere Behauptungen und deren Änderungen zurückgreift. Dabei könnte auch beobachtbar werden, wie eine zunächst namenlose Adresse mit Widersprüchen umgeht. Der konkrete Einwand betrifft deshalb die unbestimmte Zurechnungsstruktur der Befundgeschichte: Sofern wiederkehrende Urheberschaft erschließbar bleibt, kann sie unter verdeckten Namen eine eigene Reputation erzeugen. Eine Adresse erscheint dann beispielsweise als zuverlässig korrigierend oder als beharrlich ausweichend. Die angegliche Expertiseinformation erfasst diese im Verlauf zugeschriebene Kompetenz nicht notwendig; auch eine erfolgreiche Verdeckung persönlicher Identität schließt sie nicht aus. Eine positive Interaktion wäre folglich mit einer Verlagerung personenbezogenen Vertrauens auf rekonstruierte Urheberschaft vereinbar. Der randomisierte Verfahrensvergleich behält seinen Wert, während die Erklärung durch verringerte persönliche Bindungen zusätzliche Bestimmungen verlangt. Zur Schärfung sollten die Materialien ausdrücklich unterscheiden, ob nur Beziehungen zwischen Aussagen oder zusätzlich stabile Beziehungen zwischen Aussagen und Urhebern rekonstruierbar sind. Eine ergänzende Variation stabiler anonymer Urheberkennungen bei unveränderter Aussagegeschichte könnte deren Beitrag gesondert untersuchen. Ferner wären Gruppengespräche danach auszuwerten, ob eine Revision durch Belege, früheres Korrekturverhalten oder zugeschriebene Kompetenz begründet wird; solche Prozessdaten ersetzen keine kausale Mechanismenprüfung. So würde erkennbar, welche Unterscheidungen die prüfende Kommunikation selbst verwendet. Die positive Interaktion bleibt dabei eine empirische Möglichkeit; aus der systemtheoretischen Beschreibung lässt sich ihre Richtung nicht deduzieren.

Nebenlinie — sieben Kurzbegutachtungen

Hypothese. Das Selbst im Verhör (56/90).

Kernsatz. Genesis 2–3 erzählt bereits vor dem Essen sprachliche Bezugnahme auf Gebot und Handlung; die spätere Befragung fordert erstmals in dieser Erzählsequenz eine ausdrückliche Rechenschaft über die vollzogene verbotene Handlung. Als plessnersch inspirierte Lesart lässt sich dieser Übergang als öffentlich eingeforderte Selbstzuschreibung verstehen, ohne darin die Entstehung von Handlungsfähigkeit oder Selbstbewusstsein zu behaupten.

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

Separat gestartete Expertenkontexte ohne andere Gutachten. KI-Rekonstruktionen; keine Originaläußerungen der genannten Personen.

Status: Kurzbegutachtung abgeschlossen. Keine zweite Replikenrunde.

Kurzbegutachtung 1 — Kant

Diese KI-Rekonstruktion der Kant-Perspektive erkennt als Stärke die Trennung zwischen dem Vermögen zum Selbstbezug und ausdrücklich verlangter Rechenschaft. Eine Schwäche liegt darin, dass die Bestimmung des Öffentlichen als Rede gegenüber einem anderen die besondere Stellung des Fragenden nicht erfasst: In der mitgeteilten Sequenz wird Rechenschaft gegenüber der Instanz des Gebots verlangt, wodurch der Übergang einen gerichtlichen Sinn erhalten kann. Es bedarf deshalb einer Unterscheidung zwischen Selbstzuschreibung vor einem Gegenüber und Zurechnung unter einer als verbindlich gesetzten Norm. Forschungspotenzial bietet ein nach Figuren getrennter Vergleich der Deutungen anhand der Frage, welche Einzelheiten der Antworten erst durch den Begriff der Selbstzuschreibung verständlich werden. Die Prüfung von Plessner-Originalpassagen wäre hierfür ein Anschlussvorschlag, kein bereits verfügbarer Beleg.

Kurzbegutachtung 2 — Popper

Diese KI-Rekonstruktion der Popper-Perspektive sieht die Stärke in der Unterscheidung zwischen vorhandener Selbstbezüglichkeit und einer Situation, in der ausdrücklich Rechenschaft gefordert wird. Die Schwäche liegt im noch unscharfen Verhältnis von Handlungszuschreibung und Verantwortungsübernahme: Wer eine eigene Handlung einräumt und ihre Ursache einem anderen zuschreibt, übernimmt damit noch keine Schuld. Nach der vorgelegten Beschreibung wäre gerade diese Differenz am Text zu untersuchen. Mein Rat: Forderung des Fragenden, Eingeständnis einer Handlung und Übernahme oder Abwehr von Verantwortung sollten getrennt ausgelegt werden. Das Forschungspotenzial liegt darin, den zusätzlichen begrifflichen Ertrag von Selbstzuschreibung gegenüber Rechenschaft zu bestimmen; eine Gerichtsdeutung kann diese Analyse ergänzen. Ein spezifischer Plessner-Bezug bleibt bis zur Prüfung einschlägiger Originalpassagen ein Anschlussvorschlag.

Kurzbegutachtung 3 — Wittgenstein

KI-Rekonstruktion der vorgegebenen Perspektive: Die Stärke der Lesart liegt darin, dass sie nach dem Gebrauch vorhandener Fähigkeiten fragt. Doch was tut jemand, wenn er auf eine Befragung hin sein Handeln nennt? Er kann etwas einräumen und zugleich Verantwortung abweisen. Die Schwäche liegt deshalb im Sammelwort Selbstzuschreibung: Es verdeckt möglicherweise den Unterschied zwischen Eingeständnis, Rechtfertigung und Beschuldigung eines anderen. An den beschriebenen Antworten wäre satzweise zu zeigen, was jeweils zugestanden und was weitergereicht wird. Das Forschungspotenzial liegt im Vergleich dieser sprachlichen Handlungen mit der Schuld- und Gerichtslesart: Erhellte Selbstzuschreibung eine bisher übersehene Einzelheit, oder benennt sie Bekanntes um? Plessners Originaltexte wären anschließend auf einen zusätzlichen begrifflichen Gewinn zu prüfen.

Kurzbegutachtung 4 — Plessner

Diese eigenständige KI-Rekonstruktion sieht die Stärke der vorgelegten Lesart darin, dass sie eine Spannung zwischen Handeln und Sichdarstellen erschließt: Wer Auskunft gibt und Verantwortung weiterreicht, bleibt auf das eigene Tun bezogen, ohne damit schon Schuld anzuerkennen. Die Schwäche liegt darin, dass öffentlich eingeforderte Selbstzuschreibung diesen Abstand bislang eher benennt als deutet. Forschungspotenzial bietet ein figurenbezogener Vergleich von Verbergen, Selbstauskunft und Verantwortungsabwehr: Welche Formen des Selbstabstands werden jeweils leiblich vollzogen oder sprachlich artikuliert? Dabei sind Handlungsangabe und Schuldanerkennung auseinanderzuhalten und die besondere Stellung des göttlichen Fragenden zu bewahren. Erst zu prüfende Plessner-Originalpassagen wären ein philosophischer Anschluss; der hermeneutische Mehrwert müsste sich gegenüber der Schuld- und Gerichtslesart an genauer erschlossenen Textzügen zeigen.

Kurzbegutachtung 5 — Cassirer

Diese KI-Rekonstruktion der Cassirer-Perspektive sieht die Stärke der Lesart darin, Selbstzuschreibung als veränderten Gebrauch bereits vorhandener sprachlicher Möglichkeiten zu verstehen; damit bleibt Raum für unterschiedliche symbolische Gestalten des Selbstverhältnisses. Ihre Schwäche liegt in der möglichen Einebnung der religiösen Form: Gott erscheint unter dem Begriff des anderen leicht wie ein beliebiger Gesprächspartner, obwohl gerade die Eigenart seiner Anrede für die Bedeutung der Rechenschaft entscheidend sein könnte. Forschungspotenzial besitzt deshalb ein Vergleich, der an den genannten Textstationen fragt, wie göttliches Gebot, Scham und Antworten mit Verantwortungsverschiebung jeweils das Verhältnis zwischen Handelnden und Ordnung gestalten. Dabei wäre zu klären, welchen eigenen Aspekt die Rechenschaftslesart erschließt und wie dieser sich mit Schuld- und Gerichtsdeutungen verbindet.

Kurzbegutachtung 6 — Friston

Diese KI-Rekonstruktion der Friston-Perspektive sieht die Stärke der Lesart in einer Unterscheidung zwischen vorhandenem Selbstmodell und situationsabhängiger Wahl einer Antwort: Eine Figur kann ihr Handeln kennen und dennoch dessen kommunikative Darstellung verändern. Die Befragung ließe sich daher als Änderung der sozialen Anforderungen lesen. Die Schwäche liegt im Begriff Selbstzuschreibung, der Handlungszuordnung, kausale Urheberschaft und Schuldübernahme noch nicht hinreichend trennt; gerade Verantwortungsabwehr macht diese Differenz entscheidend. Das Forschungspotenzial besteht darin, die beschriebenen Antworten jeweils auf diese drei Leistungen hin auszulegen und mit einer Gerichtslesart zu vergleichen. Maßgeblich wäre der zusätzliche textliche Ertrag dieser Unterscheidung. Ein generatives Modell bliebe dabei eine Heuristik, und der Plessner-Anschluss erforderte die Prüfung einschlägiger Originalpassagen.

Kurzbegutachtung 7 — Luhmann

KI-Rekonstruktion: Die Stärke der Lesart liegt darin, Rechenschaft als kommunikative Zumutung bei bereits vorausgesetztem Selbstverhältnis zu behandeln. Die Schwäche liegt in der Bestimmung von „öffentlich“ als „gegenüber einem anderen“: Damit bleibt unterbelichtet, dass hier die Instanz fragt, auf die das verletzte Gebot zurückgeht. Diese Asymmetrie könnte die Antwortform genauer erklären als der allgemeine Bezug auf Selbstzuschreibung. Forschungspotenzial bietet deshalb eine textnahe Untersuchung der Verteilung von Fragerecht, Antwortpflicht und Verantwortungsabwehr: Welche Zuschreibung setzt die Frage voraus, welche übernimmt die Antwort, welche verschiebt sie? Die Lesart gewönne Eigenständigkeit, wenn sie an diesen Unterschieden einen Ertrag gegenüber der Gerichtsdeutung nachweist; der Plessner-Bezug bleibt bis zur Originaltextprüfung ein Anschlussvorschlag.

Quellenhinweise und offene Fragen der Nebenlinie

Die folgenden Hinweise gehören ausschließlich zu „Das Selbst im Verhör“. Sie sind ein Kandidat für spätere Quellenarbeit und wurden nicht in die Empirie-Brücke oder das Quellenstudium der Hauptlinie eingemischt.

Genesis 2–3, Einheitsübersetzung 2016. *Quelle.* Vier Textstationen am Primärtext geprüft: Gen 2,16–17; 3,2–3; 3,7–8; 3,9–13. Kein antiker Anthropogenese-Befund behauptet.

Helmuth Plessner: Die Stufen des Organischen und der Mensch. Einleitung in die philosophische Anthropologie (1928). *Quelle.* Verlagsmetadaten und Inhaltsverzeichnis geprüft. Kapitel VII Die Sphaere des Menschen als Lesekandidat; einschlägige Originalpassagen nicht gelesen, daher kein Textbeleg für Genesis-Auslegung.

Ergebnis der sieben Kurzbegutachtungen

Kant, Cassirer und Luhmann verlangen, die besondere Autorität des göttlichen Fragenden und die religiöse Gestalt der Szene gegenüber einem allgemeinen Gesprächsmodell zu bewahren. Popper und Wittgenstein unterscheiden Handlungseingeständnis, Rechtfertigung und Schuldabwehr. Plessner fragt nach dem Wechsel zwischen leiblichem Verbergen und sprachlicher Selbstdarstellung; Friston bietet ein Selbstmodell nur als Heuristik an. Keine dieser KI-Perspektiven liefert bereits einen geprüften Plessner-Originalbeleg. Der mögliche Ertrag liegt in genauerer Auslegung; eine Wiederholung der Gerichtsdeutung unter neuem Namen wäre kein Fortschritt.

Offene Prüfaufgaben

- Welchen zusätzlichen Erklärungsgewinn bietet Selbstzuschreibung gegenüber Rechenschaft und Schuldabwehr?
- Welche Originalstellen Plessners tragen die Unterscheidung zwischen einer anthropologischen Fähigkeit und ihrer situativen Ausübung?
- Kann die Gerichtslesart dieselben Textstationen mit weniger Zusatzannahmen erklären?
- Welche Rolle spielt die besondere Autorität des göttlichen Fragenden, verglichen mit einem beliebigen Gegenüber?
- Wie unterscheiden sich Handlungsangabe, kausale Urheberschaft, Schuldübernahme und Verantwortungsabwehr in jeder Antwort?

Vorläufer im Vault: Genesis - Paradies als regulatorischer Nullzustand, Reservoir - Alterität als Erstbedingung des Ich-Entwurfs 2026-05-21, Reservoir - Schuld als Strukturmerkmal reflexiver Selbst-Autorschaft 2026-05-20.

Pickup-Vorschlag: Nach Lektüre des einschlägigen Plessner-Originalkapitels die vier Genesis-Stationen anhand derselben Auslegungsfragen gegen eine Schuld- und Gerichtslesart prüfen.

Expertenrunde 2 — Repliken nur zur Hauptlinie

Jede Stimme erhielt vollständig die sechs anderen Gutachten der ersten Runde. Die Nebenlinie wurde nicht weitergeführt.

Replik 1 — Kant

Diese KI-Rekonstruktion der Kant-Perspektive bekräftigt die Unterscheidung zwischen gemessener Lösungsübereinstimmung und berechtigtem Urteil, präzisiert jedoch die Forderung meines ersten Gutachtens. Mit Wittgenstein stimme ich darin überein, dass die Kennzeichnung geltender und zurückgenommener Aussagen zur Bestimmung derselben Prüfaufgabe gehört; daraus folgt aber nicht, dass sämtliche Begründungsrelationen in beiden Bedingungen gleich zugänglich sein müssten. Eine solche Angleichung könnte gerade jene Leistung der Geschichte beseitigen, die untersucht werden soll. Gefordert ist vielmehr, dass die richtige Lösung jeweils begründbar bleibt, während der Aufwand ihrer Erschließung verschieden sein darf. Poppers Einwand erweitert meine Prüfung der Vorvalidierung: Auch ein unabhängig festgelegter Lösungsschlüssel bedarf zugänglicher Gründe und muss sachlicher Kritik offenstehen. Seine begründete Berichtigung muss sämtliche Versuchsbedingungen gleichermaßen betreffen; die ursprüngliche Auswertung bleibt gesondert auszuweisen. Luhmanns Hinweis auf Reputation ohne Namenskenntnis begrenzt schließlich die Auslegung eines Vorteils: Eine nachvollziehbare Fehlerspur kann selbst zum Träger zugeschriebener Verlässlichkeit werden. Seine Untersuchung kommunikativer Zurechnung beantwortet jedoch noch nicht die Frage, wann diese Verlässlichkeit einen berechtigten Urteilsgrund abgibt; an dieser Differenz zwischen wirksamer Orientierung und sachlicher Geltung halte ich fest. Mein Reformulierungssatz lautet: Zu prüfen ist, ob bei in jeder Darstellung begründbarer und nachprüfbarer Lösung die Zuschreibungsverzögerung mit rekonstruierbarer Geschichte einen vorab bestimmten Mindestzuwachs richtiger Korrekturen und gegenüber der Bedingung ohne rekonstruierbare Geschichte eine vorab bestimmte positive Mindestinteraktion auf der absoluten Risikodifferenzskala erreicht, ohne die vorab bestimmte Toleranz zusätzlicher Fehlkorrekturen zu überschreiten. Die ungelöste Spannung besteht darin, dass dieselbe Verknüpfung den Zugang zu Gründen erleichtern und mittelbar Ansehen erzeugen kann. Die reflektierende Urteilskraft verlangt deshalb, diese Vermittlung prüfbar zu machen, ohne aus dem Leistungsvorteil bereits die Freiheit des Urteilens von persönlicher Autorität abzuleiten.

Replik 2 — Popper

Diese KI-Rekonstruktion der Popper-Perspektive präzisiert mein erstes Gutachten: Ein kritisierbarer Lösungsschlüssel genügt erst zusammen mit einer Prüfung dessen, was die Gruppen überhaupt beurteilen können. Kant und Wittgenstein treffen einen Punkt, den ich unterschätzt habe: Nachprüfbarkeitsgarantie garantiert noch keine identische Prüfaufgabe. Daher müssen Zielbehauptung und Status zurückgezogener Angaben in allen Bedingungen kenntlich bleiben; die Lösbarkeit ist für jede Darstellung zu prüfen. Mein verblindetes Nachprüfungsverfahren für bestrittene Lösungen halte ich aufrecht, mit einheitlicher Anwendung auf sämtliche Versuchsarme und getrennter ursprünglicher und revidierter Auswertung. Friston hat ebenfalls recht: Eine Interaktion auf der Risikodifferenzskala kann aus nichtlinear umgesetzten unabhängigen Einflüssen entstehen. Sie wäre ein Verfahrensbefund, kein Nachweis einer spezifischen Verarbeitungskopplung. Seinen vorgeschlagenen zusätzlichen Modellvergleich würde ich vorerst zurückstellen: Der begrenzte Leistungstest kann scheitern, ohne dass bereits eine Prozesstheorie entscheidbar wäre. Luhmanns Einwand, nachvollziehbare Urhebererschaft könne eine namenlose Reputation tragen, verlangt eine präzisere ceteris-paribus-Klausel. Der Materialpilot muss zugeschriebene Verlässlichkeit erfassen; erkennbare Urheberbeziehungen sind zwischen den Zuschreibungsbedingungen konstant zu halten. Solche Erhebungen identifizieren den Mechanismus allerdings noch nicht. Mein Reformulierungssatz lautet: Bei in jeder Darstellungsbedingung begründbar lösbarer Aufgaben und zwischen den Zuschreibungsbedingungen gleichgehaltener Urheberstruktur erreicht verzögerte persönliche Zuschreibung die vorab festgelegten Mindestwerte für den Zuwachs der Rate richtiger Korrekturen bei rekonstruierbarer Geschichte und für die positive Interaktion auf der Risikodifferenzskala, ohne die vorab festgelegte Toleranzgrenze zusätzlicher Fehlkorrekturen zu überschreiten. Liegt bei gelungener Manipulation die obere Konfidenzgrenze eines beanspruchten Effekts unter dessen Mindestwert, scheitert die starke Kombination; breite Intervalle bestätigen sie nicht. Die Fehlkorrekturgrenze bleibt eine eigenständige Scheiterbedingung. Ungelöst bleibt die produktive Spannung, wie viel argumentativer Zusammenhang gleichgehalten werden kann, bevor die Variation der Geschichte ihren sachlichen Gehalt verliert. Dies verlangt Materialentscheidungen, keinen nachträglichen Rückzug in eine schwächere These.

Replik 3 — Wittgenstein

KI-Rekonstruktion der Wittgenstein-Perspektive: Mein erster Einwand bleibt: Dieselben Evidenzstücke sichern nicht denselben Gebrauch. Kant verschärft ihn zu Recht. Es genügt nicht, dass alle dieselbe Behauptung erkennen; die vorgegebene Lösung muss aus den jeweils zugänglichen Angaben begründbar sein. Das Wissen der Prüfer ist nicht schon das Wissen der Gruppe. Mit Popper schränke ich meine frühere Anerkennung der Vorvalidierung ein: Ihr Ergebnis muss selbst anfechtbar bleiben. Eine begründete Abweichung wird nicht durch ihre Verbuchung als Fehlkorrektur widerlegt. Für solche Einwände braucht es ein für alle Bedingungen gleiches Nachprüfungsverfahren. Luhmann zeigt eine weitere Lücke meines ersten Gutachtens. Wer sieht, dass derselbe unbekannte Urheber seine Fehler berichtigt, kann ihm vertrauen. Der Name fehlt; die Zuschreibung bleibt. Meine Forderung nach kenntlichem Geltungsstatus löst dieses Problem also nicht. Wir müssen festhalten, ob nur Behauptungen verknüpft oder auch Urheber wiedererkannt werden. Daraus folgt kein Einwand gegen den Verfahrensvergleich, wohl aber gegen seine Deutung als Befreiung von persönlicher Bindung. Fristons Einwand trifft zu: Eine Interaktion auf der Risikodifferenzskala belegt noch keine Kopplung im Inferenzprozess. Seinen Modellvergleich brauche ich für diese begriffliche Klärung jedoch nicht. Eine gut vorhergesagte Entscheidung sagt noch nicht, ob die Gruppe einen Befund berichtigt oder zunächst das Prüfmaterial versteht. Mein Reformulierungssatz lautet: Kennzeichnen Sie überall dieselbe Prüfbehauptung und den Geltungsstatus der Angaben, prüfen Sie die Begründbarkeit der Lösung für jede Darstellung und dokumentieren Sie, welche Urheberschaft trotz verdeckter Namen erkennbar bleibt. Offen bleibt eine Schwierigkeit: Frühere Selbstberichtigungen können einen guten Grund zum Vertrauen geben und zugleich die gegenwärtige Sachprüfung abkürzen. Wann nennen wir das begründetes Vertrauen, wann ungeprüfte Gefolgschaft? Auch ein richtiger Gruppenentscheid beantwortet diese Frage noch nicht.

Replik 4 — Plessner

Diese KI-Rekonstruktion der Plessner-Perspektive bekräftigt meinen ersten Einwand, erweitert aber seinen Gegenstand. Luhmann trifft genau die Schwierigkeit, dass auch eine namenlose Revisionsfolge einem Gegenüber zugerechnet werden und dessen Reputation bilden kann. Mein Vorschlag, Urheberschaft und zugeschriebene Verlässlichkeit im Materialpilot zu prüfen, bleibt daher bestehen. Seine kommunikative Adresse und die anthropologische Stellung des Urteilenden sind jedoch verschiedene Beschreibungsgegenstände: Mich interessiert, wie jemand sich im Urteil zugleich bindet und zum eigenen Urteil Abstand gewinnt. Die beobachtbare Verwendung einer Adresse erschöpft dieses Verhältnis nicht. Kant verlangt zudem mit Recht, die Zugänglichkeit der Rechtfertigungsgründe für jede Darstellungsbedingung zu prüfen. Hier muss ich meine erste Position ergänzen: Eine Geschichte kann persönliche Bindungen vermitteln und darüber hinaus die sachliche Beurteilbarkeit überhaupt erst herstellen. Prüfbehauptung und Kennzeichnung aktueller oder zurückgenommener Belege müssen deshalb in allen Bedingungen verständlich bleiben. Cassirer verschärft meinen Einwand nochmals: Schon eine Fortschrittserzählung kann Autorität gewinnen, ohne dass die Gruppe einen verlässlichen Urheber erschließt. Mein bisheriger Blick auf das vorgestellte Gegenüber war insofern zu eng; Zeitfolge, Begründungsrelation und Fortschrittsgestus sind im Material auseinanderzuhalten. Ich schlage folgenden Reformulierungssatz vor: Bei gleicher sachlicher Beurteilbarkeit und zwischen den Zuschreibungsbedingungen gleichgehaltener Revisionsurheberschaft wird geprüft, ob verzögerte persönliche Zuschreibung mit rekonstruierbarer Geschichte einen vorab bestimmten Mindestvorteil bei richtigen Korrekturen sowie gegenüber fehlender Rekonstruierbarkeit eine vorab bestimmte positive Mindestinteraktion auf der Skala absoluter Risikodifferenzen erreicht, ohne dass die Zunahme falscher Korrekturen die vorab festgelegte Toleranzgrenze überschreitet. Genauer trenne ich damit den überprüfbaren Verfahrenseffekt von der phänomenologisch zugänglichen Stellung des Urteilenden. Sachliche Nachvollziehbarkeit eröffnet möglicherweise Distanz, verbürgt sie aber nicht. Ungelöst bleibt, wie dieselbe vermittelte Beziehung Selbstkorrektur erleichtern und Bindungen festigen könnte, deren Lockerung wiederum kooperative Sorgfalt kosten kann.

Replik 5 — Cassirer

Als KI-Rekonstruktion der Cassirer-Perspektive präzisiere ich meine erste Deutung: Nachvollziehbare Verknüpfungen vermitteln nicht schon durch ihre Ordnung Geltung; sie gestalten auch, was als Grund oder Fortschritt erscheint. Kant macht dabei eine Voraussetzung sichtbar, die ich deutlicher hätte fassen müssen: In beiden Geschichtsbedingungen müssen zurückgenommene Aussagen als solche erkennbar und Lösungen aus den zugänglichen Angaben begründbar sein. Meine Unterscheidung zwischen Zeitfolge, Begründungsrelation und Fortschrittserzählung bleibt bestehen, verlangt aber zunächst diese gemeinsame Urteilsaufgabe. Mit Luhmann nehme ich ferner ernst, dass wiedererkennbare Urheberschaft auch ohne Namen Reputation erzeugen kann. Meine Rede vom Wechsel der Autoritätsträger bezeichnet daher eine Umgestaltung symbolischer Vermittlung, in der persönliche Bindungen fortbestehen können; die Kontinuität der Urheberschaft muss im Material ausdrücklich bestimmt werden. Fristons Einwand gegen die Gleichsetzung einer statistischen Interaktion mit einer Kopplung im Inferenzprozess begrenzt die Deutung eines Leistungsvorteils: Schon die nichtlineare Umsetzung unabhängiger Einflüsse kann eine solche Interaktion hervorbringen. Hier bleibt eine Differenz unserer Fragen: Ein Modell der Evidenzverarbeitung erklärt noch nicht die kulturelle Form, in der eine Revisionsfolge als Rechtfertigung verständlich wird. Mein Reformulierungssatz lautet: Bei in beiden Geschichtsbedingungen begründbarer Lösung, gleich gekennzeichnetem Aussagestatus und konstant gehaltener Urheberkontinuität wird geprüft, ob verzögerte Namensoffenlegung bei rekonstruierbarer Revisionsgeschichte die Rate richtiger Korrekturen um mindestens den vorab festgelegten Betrag steigert und die positive Interaktion auf der absoluten Risikodifferenzskala den ebenfalls vorab festgelegten Mindestwert erreicht, ohne die vorab bestimmte Toleranzgrenze zusätzlicher Fehlkorrekturen zu überschreiten. Ungelöst bleibt die produktive Spannung, dass jene Geschichte, die Gründe zugänglich macht, durch dieselbe Gestaltung Vertrauen in eine namenlose Instanz bilden kann. Sie lässt sich nicht allein durch bessere Verblindung beseitigen; mein ursprünglicher Vorschlag, die Darstellung von Fortschritt gesondert zu variieren, erhält dadurch zusätzliches Gewicht.

Replik 6 — Friston

Diese KI-Rekonstruktion erweitert meine erste Position: Die Revisionsgeschichte kann außer Annahmen über Evidenzabhängigkeiten auch die Präzision sozialer Erwartungen verändern. Luhmanns Einwand trifft deshalb meine zunächst zu saubere Zuordnung der beiden Eingriffe: Verdeckte Namen verhindern keine Reputation, wenn wiederkehrende Urheberschaft erschließbar bleibt. Das generative model muss eine aus dem Korrekturverhalten erschlossene Quellenverlässlichkeit zulassen; eine erfolgreiche Verblindung schließt diesen Mechanismus nicht aus. Cassirers Unterscheidung von Zeitfolge, Begründungsrelation und Fortschrittserzählung verlangt außerdem, argumentative Verknüpfung von der suggestiven Autorität späterer Fassungen im Material zu trennen. Mit Kant stimme ich darin überein, dass die Lösung unter jeder Darstellung begründbar bleiben muss und gleiche Evidenzmengen dies nicht gewährleisten. Sein Rechtfertigungsanspruch geht allerdings über meine Erklärung der Urteilsbildung hinaus: Eine hohe posterior probability verbürgt keine normative Berechtigung; umgekehrt ersetzt eine tragfähige Begründung kein Modell ihrer tatsächlichen Nutzung. Meinen ursprünglichen Einwand bekräftige ich: Eine positive Interaktion auf der Risikodifferenzskala kann schon bei additiven Effekten auf die Log-Odds entstehen. Sie allein identifiziert daher keine Kopplung im Inferenzprozess, und das free energy principle bestimmt ihr Vorzeichen nicht. Vorab spezifizierte Modelle unabhängiger Beiträge und gekoppelter Prozesse sollten anhand individueller und gemeinsamer Wahrscheinlichkeitsurteile verglichen werden; Simulationen müssen zuvor zeigen, ob Evidenzgewichtung, Quellenverlässlichkeit und Entscheidungsschwelle unterscheidbar sind. Mein Reformulierungssatz lautet: Bei überall erkennbarer Prüfbehauptung und erkennbarem Rücknahmestatus werden der vorab festgelegte Mindestnutzen verzögerter Zuschreibung bei rekonstruierbarer Geschichte, die positive Mindestinteraktion auf der Risikodifferenzskala und die Fehlkorrekturgrenze geprüft; ergänzende Prozessmodelle berücksichtigen, dass Geschichte sowohl Evidenzabhängigkeiten erschließen als auch Reputation erzeugen kann. Die produktive Spannung bleibt, dass dieselbe Geschichte sachliche Prüfung erleichtern und personenbezogenes Vertrauen neu organisieren kann; dieses Vertrauen vollständig zu entfernen könnte gerade den untersuchten Gegenstand verändern.

Replik 7 — Luhmann

KI-Rekonstruktion: Meine erste Stellungnahme zur namenlosen Reputation bleibt bestehen, bedarf jedoch einer Ergänzung: Auch ohne stabile Urheberadresse kann die Darstellung Vertrauen organisieren. Cassirers Unterscheidung von Zeitfolge, Begründungsrelation und Fortschrittserzählung zeigt, dass auch die Anordnung einer Geschichte Autorität vermitteln kann. Das Spätere erscheint dann womöglich als das Bessere, bevor seine Gründe geprüft werden. Zu beobachten ist deshalb, woran Kommunikation Vertrauen anschließt: an zugerechnetes Korrekturverhalten oder an eine Darstellung, die Bewährung suggeriert. Mit Plessner besteht hier Konvergenz; seine anthropologische Deutung übernehme ich nicht. Die Verlässlichkeit des Gegenübers wird hier als kommunikative Erwartungsstruktur untersucht. Kant und Wittgenstein nötigen außerdem zur Präzisierung einer Voraussetzung meines Vorschlags: Ohne übereinstimmend erkennbaren Status aktueller und zurückgezogener Aussagen könnte die Geschichte erst bestimmen, was überhaupt geprüft wird. Gleiche Evidenzmenge genügt also nicht. Die Prüfbehauptung und diese Statusangaben müssen in sämtlichen Bedingungen erkennbar bleiben. Fristons Einwand akzeptiere ich: Die Interaktion auf der Risikodifferenzskala belegt keine Kopplung der zugrunde gelegten Inferenzprozesse. Schon eine nichtlineare Übersetzung unabhängiger Beiträge in Entscheidungen kann einen solchen Skalenbefund erzeugen. Ein solches Prozessmodell müsste zusätzlich erläutern, wie individuelle Urteile mit kommunikativen Anschlüssen zusammenhängen; die Gruppe ist kein vergrößertes Bewusstsein. Meine Zurückhaltung gegenüber einer theoretischen Vorzeichenvorhersage bleibt bestehen. Der konkrete Reformulierungssatz lautet: Bei konstantem Prüfgegenstand und Aussagestatus sowie explizit festgelegter Urheberzurechnung wird geprüft, ob bis zum fixierten Ersturteil verzögerte Namensoffenlegung bei rekonstruierbarer Geschichte den vorab bestimmten Mindestzuwachs der Rate richtiger Korrekturen und die minimale positive Interaktion auf der Risikodifferenzskala erreicht, ohne die vorab festgelegte Toleranz für zusätzliche falsche Korrekturen zu überschreiten. Die Spannung bleibt ungelöst: Dieselbe Rekonstruierbarkeit kann die Prüfung von Gründen ermöglichen und Erwartungen stabilisieren, die weitere Prüfung entbehrlich erscheinen lassen; ob daraus bessere Korrekturen entstehen, entscheidet sich nicht schon durch die Beschreibung dieser Möglichkeit.

Synthese im Sokrates-Modus

Konvergenzen

Wenn ich die vierzehn Gutachten recht verstehe, so stimmen Kant, Popper und Wittgenstein darin überein, dass dieselbe Menge an Belegen noch keine vergleichbare Prüfaufgabe ergibt. In jeder Darstellung müssen die gegenwärtige Behauptung, zurückgezogene Angaben und eine begründbare Lösung erkennbar bleiben. Cassirer und Luhmann schließen sich dieser Forderung ausdrücklich an. Plessner, Luhmann und Cassirer treffen sich ferner in einer unbequemen Einsicht: Auch ohne Namen kann eine Geschichte Vertrauen in ihren Urheber oder in ihre eigene Ordnung erzeugen. Was verbirgt die Namensverdeckung dann tatsächlich? Zunächst nur eine bestimmte persönliche Information. Friston, Popper und Wittgenstein verlangen schließlich, einen gemessenen Zusatznutzen von seiner Erklärung zu unterscheiden. Eine statistische Interaktion beweist weder die Überwindung persönlicher Autorität noch eine besondere Kopplung geistiger Prozesse. Diese Übereinstimmungen schärfen die Bedingungen eines Versuchs; sie besagen nicht, dass sein erwartetes Ergebnis eintreten wird.

Divergenzen

Wo widersprechen sich die Stimmen unmittelbar? Die deutlichste methodische Differenz besteht zwischen Friston und Popper: Friston fordert ergänzende Prozessmodelle und deren Identifizierbarkeit, während Popper diesen Vergleich für den ersten begrenzten Leistungstest zurückstellen möchte. Wir können den Verfahrensvergleich deshalb primär prüfen und Mechanismen als spätere Untersuchung kennzeichnen. Eine begriffliche Differenz liegt zwischen Kants berechtigtem Urteil und Luhmanns beobachtbarer Vertrauenskommunikation. Beide Fragen sind zulässig, doch ein brauchbarer Kommunikationsanschluss beantwortet nicht schon die Frage nach guten Gründen. Plessners Stellung des Urteilenden lässt sich ebenso wenig mit einer stabilen Kommunikationsadresse gleichsetzen. Eine substantiell entgegengesetzte Vorhersage über das Vorzeichen des Effekts tragen die Repliken dagegen nicht aus. Sie lassen offen, ob Namenskenntnis Prüfung hemmt oder durch Verantwortlichkeit verbessert. Einen solchen Streit künstlich zwischen den Personen zu inszenieren wäre unredlich. Es bleibt eine gemeinsame, empirisch ungelöste Gegenmöglichkeit.

Produktive Antinomien

Welche Spannung würde verschwinden, wenn wir die These zu bequem formulierten? Bei Cassirer und Luhmann macht die Revisionsgeschichte Gründe zugänglich und kann zugleich selbst zur Autorität werden. Eine Gruppe könnte besser urteilen, weil sie nachprüft, oder weil sie einer als bewährt erscheinenden Fehlerspur vertraut. Kant erinnert daran, dass diese Wege nicht dieselbe Rechtfertigungsleistung besitzen. Die zweite Spannung verbindet Plessners Abstand zum eigenen Urteil mit der von Popper offengehaltenen Gegenrichtung: Persönliche Bindung kann sachliche Korrektur erschweren, aber auch die kooperative Sorgfalt fördern, die eine Korrektur erst ermöglicht. Beide Spannungen sind produktive Gegenrichtungen, keine logischen Antinomien im strengen kantischen Sinn. Wir halten sie fest, indem wir richtige und falsche Änderungen getrennt messen und die Gründe der Gruppenentscheidung ergänzend dokumentieren. Wir dürfen sie nicht nutzen, um jedes beliebige Ergebnis zur Bestätigung derselben starken Behauptung umzudeuten.

Reformulierungs-Anstoß

Lass uns Poppers Reformulierung als Ausgangspunkt nehmen und fragen, was Kant, Wittgenstein und Luhmann ihr hinzufügen. Sie verlangt einen positiven Mindestnutzen, eine positive Mindestinteraktion und eine eigenständige Grenze für Fehlerkorrekturen. Kant und Wittgenstein sichern die Begründbarkeit der Lösung in jeder Darstellung; Luhmann verlangt eine ausdrücklich festgelegte Urheberzurechnung. Cassirer schützt vor einer Fortschrittserzählung, in der später schon besser bedeutet. Die stärkste überprüfbare Fassung betrifft damit zunächst nachrechenbare Methodenfehler und zuvor unabhängige Prüfgruppen, nicht die gesamte wissenschaftliche Wahrheitssuche. Rekonstruierbare Geschichte darf den Aufwand des Nachvollziehens verändern, ohne nur einem Versuchsarm entscheidende Tatsachen zugänglich zu machen. Fristons Prozessmodelle bleiben eine mögliche Anschlussprüfung. Vor der Datenerhebung müssen Materialien, Zielgruppe und Entscheidungsgrenzen feststehen. Die folgende quantitative Fassung setzt solche Grenzen ausdrücklich als prüfbar Entwurf; sie sind weder aus einer Theorie deduziert noch bereits empirisch gerechtfertigt.

Offene Frage

Welche Frage müssen wir als Nächstes beantworten, damit dieser Entwurf mehr als ein sauber klingendes Verfahren wird? Wittgensteins Einwand lenkt auf die Herstellung der Materialien: Lassen sich dieselben gültigen Angaben so darstellen, dass eine Lösung überall begründbar bleibt und nur ihre Revisionsbeziehungen unterschiedlich gut rekonstruierbar sind? Cassirer würde ergänzen, dass dabei keine Fortschrittserzählung heimlich den Lösungshinweis übernimmt. Luhmann verlangt ferner eine dokumentierte Urheberstruktur, damit Namensverdeckung nicht mit dem Verschwinden aller sozialen Zurechnung verwechselt wird. Deshalb soll der nächste Pickup eine konkrete nachrechenbare Aufgabe samt vier Darstellungen entwickeln und unabhängig prüfen lassen. Erst danach lässt sich entscheiden, ob der geplante Vergleich die behauptete Sache trifft. Plessners Frage nach der erlebten Bindung bleibt als spätere Erweiterung erhalten. Der neue Verzweigungstag benennt zunächst die dringendere Materialfrage, an der ein ehrlicher Hauptversuch beginnen oder schon vorab scheitern kann.

Finale Hypothese

Kernsatz. Bei unabhängig zusammengesetzten Prüfgruppen und in jeder Darstellung begründbar lösbaren Methodenaufgaben erhöht das Zurückhalten der Autoridentität bis zur Fixierung des gemeinsamen Ersturteils die Rate richtiger Korrekturen bei rekonstruierbarer Revisionsgeschichte gegenüber früher Offenlegung um mindestens fünf Prozentpunkte. Dieser Vorteil liegt auf der Skala absoluter Risikodifferenzen mindestens drei Prozentpunkte über dem entsprechenden Vorteil ohne rekonstruierbare Revisionsbeziehungen, während bei vorhandener Geschichte die Rate falscher Änderungen an richtigen Ausgangsbefunden um höchstens zwei Prozentpunkte steigt.

(a) Strukturthese. Sachliche Revisionsbeziehungen und persönliche Zuschreibungen sind unterscheidbare Informationsangebote. Namenlose Darstellungen können weiterhin Reputation und narrative Autorität erzeugen; die Unterscheidung begründet keine allgemeine Rangordnung beider Angebote.

(b) Empiriethese. Bei unabhängig zusammengesetzten Prüfgruppen und in jeder Darstellung begründbar lösbaren Methodenaufgaben erhöht das Zurückhalten der Autoridentität bis zur Fixierung des gemeinsamen Ersturteils die Rate richtiger Korrekturen bei rekonstruierbarer Revisionsgeschichte gegenüber früher Offenlegung um mindestens fünf Prozentpunkte. Dieser Vorteil liegt auf der Skala absoluter Risikodifferenzen mindestens drei Prozentpunkte über dem entsprechenden Vorteil ohne rekonstruierbare Revisionsbeziehungen, während bei vorhandener Geschichte die Rate falscher Änderungen an richtigen Ausgangsbefunden um höchstens zwei Prozentpunkte steigt.

Begründung. Luhmann - Vertrauen als Komplexitätsreduktion eröffnet die Frage, worauf eine Gruppe Verlässlichkeit stützt; daraus folgt kein festes Vorzeichen eines Namenseffekts. Die Provenienzfrage aus 06 Hypothesentag/2026-10-05 wird auf die nachvollziehbare Änderung von Befunden übertragen. Kant und Wittgenstein verlangen gleiche Prüfbehauptungen und erkennbare Geltungsstände, Popper einen anfechtbaren Lösungsschlüssel. Cassirer und Plessner begrenzen die Deutung, weil eine namenlose Geschichte persönliche oder symbolische Bindungen erhalten kann.

Falsifikationsbedingung. Bei vorab festgelegten, bestandenen Material- und Manipulationsprüfungen scheitert diese quantitative Fassung im registrierten Anwendungsbereich, wenn die obere simultane 95-Prozent-Konfidenzgrenze des Verzögerungsvorteils bei vorhandener Geschichte unter fünf Prozentpunkten oder diejenige der Interaktion unter drei Prozentpunkten liegt, oder wenn die untere Grenze der Zunahme falscher Korrekturen bei vorhandener Geschichte über zwei Prozentpunkten liegt. Breite Intervalle, die weder Bestehen noch Scheitern erlauben, lassen den Versuch unentschieden; eine misslungene Manipulation erlaubt keinen Sachentscheid über die These.

Geltungsbereich. Planungsentwurf für unabhängig zusammengesetzte Dreiergruppen mit überprüften Grundkenntnissen empirischer Methoden; mehrere nachrechenbare oder reproduzierbare Methodenaufgaben, jeweils mit vorab festgelegter Zielbehauptung. Keine Verallgemeinerung auf offene Forschungsfragen oder bestehende Laborteams.

Versuchsaufbau. Gruppenrandomisiertes 2×2-Design: rekonstruierbare versus nicht verknüpfte Revisionsbeziehungen; persönliche Identität früh versus erst nach fixiertem Ersturteil offen. Gleiche Belegstücke, erkennbare Geltungsstände, Zeitbudgets, relevante Expertiseangaben und Revisionsurheberstruktur innerhalb beider Namensbedingungen. Der Aufwand des Erschließens darf sich zwischen den Geschichtsbedingungen unterscheiden. Fehlerhafte und richtige Ausgangsbefunde werden getrennt und bei allen zugewiesenen Gruppen ausgewertet.

Schwellenstatus. Fünf und drei Prozentpunkte Mindestnutzen sowie zwei Prozentpunkte Fehlerkorrekturtoleranz sind gesetzte Grenzen dieser starken Prüffassung. Sie sind keine Literaturbefunde und noch nicht durch Anwendungskosten begründet. Vor dem Hauptversuch sind ihre praktische Angemessenheit, Materialien und Präzisionsplanung unabhängig von dessen Ergebnissen zu begründen und zu registrieren. Eine Änderung würde eine neue quantitative Fassung schaffen, nicht die vorliegende nachträglich retten.

Deutungsgrenze. Eine Interaktion auf der Risikodifferenzskala ist ein Verfahrensbefund. Sie kann bei auf anderer Skala unabhängigen Einflüssen auftreten und belegt weder einen Statusmechanismus noch eine Kopplung interner Inferenzprozesse.

Finale Bewertung mit Begründung

Interne Schlussbewertung vor externer Prüfung, Skala 1–10.

Kriterium	Score	Begründung
Originalität	6	Als möglicher Eigenbeitrag verbleibt der eng definierte Interaktionsvergleich; verzögerte Namensoffenlegung und Revisionsverläufe sind bereits bekannt.
Falsifizierbarkeit	8	Drei getrennte Entscheidungsgrößen und ihre Scheitergrenzen sind bestimmt; Materialvalidierung und Präzisionsplanung fehlen noch.
Begriffliche Klarheit	8	Prüfgegenstand, Geltungsstand, Urheberstruktur und Ergebnisgrößen sind getrennt; ihre gelungene Umsetzung muss ein Pilot zeigen.
Tiefe	7	Das Verhältnis von Gründen, Vertrauen und Autorität wird sichtbar, während der Versuch zunächst nur einen begrenzten Leistungseffekt prüft.
Forschungsrelevanz	8	Die These schließt an experimentelle Forschung zu Autoridentität, Begutachtung und Fehlererkennung an.
Interdisziplinäre Anschlussfähigkeit	7	Metawissenschaft, Sozialpsychologie und Informationsgestaltung haben konkrete Anschlüsse, deren Reichweite durch die eingegrenzten Aufgaben begrenzt bleibt.
Vault-Anschluss	5	Vertrauen und die vorige Provenienzfürage liefern Anschlussstellen, aber keine ausgearbeitete Herleitung des behaupteten Effekts.
Antinomie-Test	7	Die plausible Gegenthese eines Vorteils früher Verantwortungszurechnung bleibt am selben Korrekturergebnis prüfbar.
Publikationsmöglichkeit	7	Ein empirischer Beitrag zur Begutachtungsforschung ist denkbar; Pilot, registrierter Versuchsplan und Ergebnisse stehen aus.
Summe, gewichtet auf 90	62	Gewichtssumme 10,0; Normierung 0,9.

Lerneffekt der Pipeline

Erstbewertung **63/90**, interne Schlussbewertung **62/90**, Differenz **-1**. Interdisziplinäre Anschlussfähigkeit sinkt von 8 auf 7, weil der erste Versuch auf nachprüfbare Methodenaufgaben begrenzt wird. Die übrigen Scores bleiben gleich: begriffliche Verbesserungen beseitigen die offenen Material- und Validierungsfragen nicht. Der Nutzen des Panels besteht in überprüfbaren Einschränkungen, nicht in einer automatischen Punkteerhöhung.

Frage an die nächste Runde

#verzweigung-offen-fehlerspur-materialvergleich — Wie lassen sich gültige Angaben gleich zugänglich halten und nur ihre Revisionsbeziehungen variieren, ohne den Gruppen unterschiedliche Aufgaben zu stellen?

Empfohlener Pickup-Anlass. Sobald eine konkrete Methodenaufgabe mit dokumentierter Revisionsgeschichte für einen unabhängigen Materialpilot vorliegt.

Anschlussverbindungen. Luhmann - Vertrauen als Komplexitätsreduktion, 06
Hypothesentag/2026-10-05.

Qualitätsmodus. Ziel und verifiziertes Sitzungsmodell: GPT-6 Astra, Reasoning max; getrennte Panelaufufe explizit entsprechend konfiguriert. Die CLI-Belege dokumentieren erfolgreiche Abschlüsse und Kontexte, keine zusätzliche unabhängige Modellattestierung des Providers.

Präzisierung nach externer Prüfung

Korrektur bedeutet die Änderung des vorgelegten Ausgangsbefunds durch das fixierte gemeinsame Ersturteil. Es gibt davor kein eigenes Gruppenurteil, auf dessen Fehler konditioniert wird. Im späten Arm bleibt die Identität bis zu diesem Zeitpunkt verdeckt. Die nachfolgende Offenlegung und erneute Entscheidungen gehören zu einem gesonderten Folgeversuch; ihre Wirkung wird hier nicht behauptet.

Die schul-fremde Prüfung bestätigt die notwendige Trennung von vorgelegtem Befund und eigenem Gruppenurteil; sie wird ausdrücklich festgeschrieben. Darüber hinaus umfasst der mögliche Verfahrenseffekt auch unterschiedlich verbleibende Prüfzeit und Erschließungsaufwand, nicht ausschließlich ein besseres Verständnis von Revisionsbeziehungen. Bekannte, unbekannte und umstrittene Autoren dürfen nicht als austauschbare Identitätshinweise gelten: Die Profile und ihre Vertrautheit gehören zur registrierten Material- und Zielgruppendefinition. Ein Bestehen der drei Schwellen garantiert keine bessere gesamte Endgenauigkeit oder positive Nutzenbilanz; dafür sind die Fehlerhäufigkeit der konkreten Anwendung und die Kosten falscher Änderungen entscheidend. Der Test hält Aufgaben und Richtigkeitsstandards fest und untersucht deshalb keine historische Veränderung solcher Standards. Als eigene spätere Untersuchung bleibt offen, ob sichtbar gemachte Korrekturraten und die Einstufung als verlässlicher Prüfer das Urteil und die Dokumentation selbst verändern. Solche Rückwirkungen werden als prüfbare Möglichkeiten geführt, nicht mit dem Begriff *looping effects* schon bewiesen.

Die interne Tabelle bleibt als vorheriger Bewertungsstand erhalten. Die externe Schlussbewertung beträgt **61/90**; Originalität sinkt von 6 auf 5. Alle neun externen Scores mit Begründung folgen im Prüfungsanhang.

Empirie-Brücke (Phase 3.5, GPT-6 Astra mit Websuche)

Die folgenden Konsequenzen stammen aus der quantitativen Empiriethese. Die begriffliche Unterscheidung von Gründen und Zuschreibungen liefert für sich keine Messwerte. Aktuelle Recherche nach der Synthese am 13.10.2026; keine eigenen Daten erhoben.

Empirische Konsequenzen

1. **Positiver Korrekturvorteil.** Bei rekonstruierbarer Geschichte muss spätere Autoridentifikation mindestens fünf Prozentpunkte mehr richtige Korrekturen ermöglichen. Beobachtbar in der fixierten Erstentscheidung zufällig zugeteilter Prüfgruppen über fehlerhafte Methodenaufgaben.
2. **Zusatznutzen der Kombination.** Der Verzögerungsvorteil muss mit rekonstruierbarer Geschichte mindestens drei Prozentpunkte größer ausfallen als ohne verknüpfte Revisionsbeziehungen. Beobachtbar als Interaktion auf der absoluten Risikodifferenzskala im Viergruppenvergleich.
3. **Begrenzte Fehlkorrekturen.** Der Vorteil darf bei vorhandener Geschichte mit höchstens zwei Prozentpunkten zusätzlicher falscher Änderungen erkaufte werden. Beobachtbar an gesonderten, ursprünglich richtigen Aufgaben derselben Gruppen.

Bestehende Befunde

Zu Konsequenz 1

Stand: offen; benachbarte Befunde gemischt. Shah (2023) dokumentiert bereits eine verzögerte Autoroffenlegung, ohne unseren Richtigkeitsvergleich. Huber et al. (2022) zeigen Statuswirkungen; daraus folgt kein Mindestgewinn sachlicher Korrektur. Godlee, Gale und Martyn (1998) fanden bei eingebauten Fehlern keinen signifikanten Erkennungsvorteil der untersuchten Verblindung. Die starke Richtungsbehauptung bleibt daher eine Hypothese.

Zu Konsequenz 2

Stand: offen. Rastogi et al. (2024) untersuchen Namenskenntnis zwischen diskutierenden Gutachtern, jedoch keine Autorverzögerung oder Revisionsgeschichte. Singh et al. (2026) berichten beim Code Review keine statistisch gestützte Wechselwirkung von Komplexität und Mensch/KI-Herkunftslabel. Das ist ein angrenzender Vergleich mit anderen Faktoren, kein direkter Gegenbeweis. Keine geprüfte Quelle testet unsere genau bezeichnete Interaktion.

Zu Konsequenz 3

Stand: offen. Singh et al. trennen die Annahme richtiger und falscher Revisionsvorschläge und liefern damit einen methodischen Anschluss, aber keinen Nachweis unserer Zwei-Prozentpunkte-Grenze. Liu et al. (2024) fanden in ihrem Ankerexperiment keinen signifikanten Beharrungseffekt; daraus folgt keine allgemeine Fehlerfreiheit. Leek, Taub und Pineda (2011) motivieren als Gegenrichtung den möglichen Nutzen von Kooperation und Verantwortung.

Riskante Vorhersage (Schwellentest)

Vorhersage. $B = p(\text{richtige Korrektur} \mid \text{Geschichte, spät}) - p(\text{richtige Korrektur} \mid \text{Geschichte, früh}) \geq 0,05$.
 $I = B - [p(\text{richtige Korrektur} \mid \text{ohne verknüpfte Geschichte, spät}) - p(\text{richtige Korrektur} \mid \text{ohne verknüpfte Geschichte, früh})] \geq 0,03$. $F = p(\text{falsche Änderung} \mid \text{Geschichte, spät}) - p(\text{falsche Änderung} \mid \text{Geschichte, früh}) \leq 0,02$. Alle drei Bedingungen sind notwendig.

Status und Begründung der Grenzen. Die Werte 0,05, 0,03 und 0,02 legen eine starke, quantitativ angreifbare Prüffassung fest. Sie sind bewusst gesetzte Planungsschwellen, keine aus der Literatur geschätzten Effekte. Ob der Nutzen fünf zusätzlicher richtiger Korrekturen je hundert fehlerhafter Aufgaben eine Zunahme um höchstens zwei falsche Änderungen je hundert richtiger Aufgaben rechtfertigt, hängt von Fehlerraten und Schadensgewichten ab. Der Interaktionswert bezeichnet einen Unterschied zwischen Effekten, keinen einfachen Zusatz pro hundert Personen. Vor einem Hauptversuch müssen Anwendungsbegründung und Präzisionsplanung feststehen; ein nachträglicher Schwellenwechsel rettet diese Fassung nicht.

Methodenvorschlag. Unabhängig zusammengesetzte Dreiergruppen mit vorab überprüfter Methodenkompetenz werden zufällig einem von vier Armen zugeteilt. Jede Gruppe prüft mehrere vorab validierte nachrechenbare oder reproduzierbare Methodenaufgaben. Ein festes, in allen Armen gleich gewichtetes Set enthält fehlerhafte und richtige Ausgangsbefunde; pro Aufgabe ist genau eine gegenwärtige Zielbehauptung zu beurteilen. Geschichte bedeutet verknüpfte Änderungsschritte, der Vergleich dieselben Belege und Geltungsmarkierungen ohne diese Verknüpfungen. Früh/spät betrifft Namen und persönliche Statushinweise derselben Autoren; aufgabenrelevante Expertiseangaben bleiben gleich. Urheberkontinuität, spätere Verantwortlichkeit, Zeit und Kommunikation zwischen Prüfgruppen werden kontrolliert. Ein gesperrtes gemeinsames Ersturteil ist der Messzeitpunkt, erst danach erfolgt im späten Arm die Offenlegung.

Vorabprüfung. Vor dem Hauptversuch prüft eine andere Stichprobe Erkennbarkeit der Zielbehauptung, aktuelle und zurückgezogene Angaben, sachliche Lösbarkeit, Rekonstruierbarkeit der Revisionen sowie erkennbare Autoridentität und Urheberkontinuität. Die Schwellen für diese Prüfungen und der zulässige Anteil durchgesickerter Identitäten sind noch zu bestimmen und müssen vor Ergebniseinsicht fixiert sein. Lösbarkeit bedeutet zugängliche Gründe, nicht gleich hohe Lösungsraten; sonst würde ein möglicher Verfahrenseffekt wegkalibriert. Narrative Wertungen wie neueste gleich beste Fassung werden vermieden.

Auswertung und Präzision. Die primären Raten werden über alle zugewiesenen Gruppen und die feste Aufgabenmischung standardisiert. Richtige Korrekturen haben alle fehlerhaften, falsche Änderungen alle

richtigen zugewiesenen Aufgaben als Nenner. Gruppen- und Aufgabenabhängigkeiten sowie fehlende Entscheidungen werden vorab modelliert; fehlende Angaben dürfen die Nenner nicht ergebnisabhängig verkleinern. Ein clusterberücksichtigendes Verfahren erzeugt simultane 95-Prozent-Konfidenzintervalle für B, I und F auf der Risikodifferenzskala. Fallzahl und konkrete Intervallmethode werden vorab anhand plausibler Basisraten, Gruppen- und Aufgabenstreuung simuliert; hier wurde weder Power berechnet noch ein Versuch durchgeführt.

Was wäre der widerlegende Befund? Vorläufiges Bestehen nur, wenn untere Grenzen für B und I mindestens 0,05 beziehungsweise 0,03 betragen und die obere Grenze für F höchstens 0,02 ist. Scheitern, wenn eine obere Grenze für B oder I unter ihrer Mindestschwelle liegt oder die untere Grenze für F über 0,02 liegt. Andernfalls unentschieden. Die Regel bezieht sich auf simultane Intervalle und vorausgesetzte gültige Manipulationen, nicht auf einzelne p-Werte.

Anfechtbare Referenz. Unabhängige Prüfer begründen den Lösungsschlüssel durch Rechnung oder Reproduktion. Ein vorab beschriebenes, zur Bedingung verblindetes Einspruchsverfahren prüft strittige Fälle einheitlich für alle Arme. ursprüngliche und gegebenenfalls korrigierte Auswertung bleiben getrennt sichtbar. Eine freie nachträgliche Umklassifizierung ungünstiger Fälle ist ausgeschlossen.

Deutungsgrenze. Begründungstexte und Vertrauensurteile können ergänzend beschreiben, wie Gruppen entscheiden. Sie dürfen im primären Vergleich nicht als nachträgliche Ausschlusskriterien oder wegkontrollierte Vermittler dienen. Prozessmodelle und eine getrennte Variation narrativer Rahmung wären weitere Experimente; der Primärtest allein identifiziert sie nicht.

Daten- und Materiallage

Das öffentliche [Repository](#) zu Liu et al. enthält Material und anonymisierte Wertungen; es wurde nicht neu ausgewertet. [ARIES, D'Arcy et al. \(2024\)](#) bietet zugeordnete Gutachterkommentare und Textänderungen als möglichen Materialbestand. Beide enthalten keinen fertigen Datensatz für den vorgeschlagenen Viergruppenversuch.

Offene empirische Fragen

- #verzweigung-offen-empirie-fehlerspur-material — Vier faire Darstellungen einer konkret nachrechenbaren Aufgabe entwickeln und unabhängig validieren.
- #verzweigung-offen-empirie-fehlerspur-schwellen — Mindestnutzen, Fehlkorrekturtoleranz und Präzisionsplanung anhand einer konkreten Anwendung begründen.
- #verzweigung-offen-empirie-fehlerspur-reputation — Begründungsnutzung von namenloser Reputation und Fortschrittsrahmung unterscheiden.

Empirie-Score

Score: 7/10. Konsequenzen und Entscheidungsgrößen sind ableitbar, benachbarte Versuche und Materialien verfügbar. Der konkrete Aufgabenbestand, die Anwendungsbegründung der Grenzen und die simulationsbasierte Präzisionsplanung fehlen. Der Score bezeichnet einen konstruierbaren Prüfpfad, keine Bestätigungswahrscheinlichkeit.

Präzisierung nach dem Falsifikations-Audit

Der Audit bestätigt die drei notwendigen Entscheidungsbedingungen und warnt zu Recht vor wechselnden Nennern, nachträglichen Manipulationsausreden und ergebnisabhängigen Schwellen. Seine Forderung, die primäre Korrektur erst nach der späten Namensoffenlegung zu messen, wird dagegen nicht übernommen: Das würde den Zielparameter wechseln. Gemeint ist die Korrektur eines vorgelegten Ausgangsbefunds durch das verbindliche gemeinsame Ersturteil; die späte Gruppe urteilt zu diesem Zeitpunkt bei zurückgehaltener Identität. Die spätere Offenlegung verursacht nicht rückwirkend die gemessene Entscheidung. Die finale Formulierung bezeichnet deshalb ausdrücklich das Zurückhalten der Identität. Ein anschließendes Revisionsfenster wäre ein eigener Folgeversuch. Es wird nicht auf anfänglich falsche eigene Gruppenurteile konditioniert. Übernommen werden eine konkrete Defaultregel für nicht eingereichte Änderungen, Sensitivitätsgrenzen bei technisch unbestimmbaren Endzuständen sowie die Präzisionsplanung mit Abständen zu den beanspruchten Grenzwerten. Alle Änderungen sind Klarstellungen; die drei numerischen Schwellen bleiben unverändert.

Am festgelegten Fristende gilt keine eingereichte Änderung als Beibehaltung des Ausgangsbefunds: bei fehlerhaften Aufgaben keine richtige Korrektur, bei richtigen Aufgaben keine falsche Änderung. Diese Defaultregel muss in Material und Software gleich umgesetzt werden. Tatsächlich technisch unbestimmbare Endzustände werden zusätzlich durch vorab festgelegte Best-/Worst-Case-Sensitivitätsgrenzen behandelt; ein Bestehen wird nur behauptet, wenn es auch unter den für die These ungünstigen Zuordnungen erhalten bleibt.

Die Simulation muss wahre Effektgrößen mit verschiedenen Abständen zu den drei Entscheidungsgrenzen betrachten. Genau an einer Grenze ist eine hohe Bestehenswahrscheinlichkeit durch bloße Fallzahlerhöhung nicht zu erwarten.

Externe Begutachtung

Die Originalitätsprüfung nutzt aktuelle Primärliteratur. Falsifikations- und schul-fremde Prüfung liefen in getrennten Codex-Kontexten mit GPT-6 Astra/max. Die schul-fremde Stimme erhielt nur Hypothese und begrenzten Synthesekontext, keine Empirie-Vorbefunde und keine anderen externen Texte. Gleiche Modellfamilie bedeutet eine verbleibende Grenze der Unabhängigkeit. Die Perspektiven sind KI-Rekonstruktionen.

Stage 1 — Originalitätsprüfung (Websuche)

Anschlussfähigkeit (was ist bekannt)

Die begriffliche Trennung von persönlicher Identität und sichtbar gemachter Inhaltsgeschichte ist vorweggenommen: Stuart et al. (2012) unterscheiden Identitäts-, Inhalts- und Interaktionstransparenz und behandeln dabei ausdrücklich Revisionsverläufe. Das verhindert, die Strukturthese als begriffliche Erstentdeckung auszugeben. Shah (2023) dokumentiert zudem bereits verzögerte Namensoffenlegung im wissenschaftlichen Begutachten. Die technische Verbindung von Revisionsansicht und verdeckten Identifikationsdateien ist in der offiziellen Digital-Commons-Dokumentation beschrieben; sie ist kein Wirksamkeitsnachweis.

Auch die umgebenden empirischen Probleme sind etabliert. Rastogi et al. (2024) untersuchen die Sichtbarkeit von Gutachteridentitäten in gemeinsamen Diskussionen. Singh et al. (2026) kombinieren Codekomplexität und Herkunftslabel und unterscheiden angemessene von falschen Revisionsannahmen. Diese Faktoren sind nicht identisch mit unseren; deshalb widerlegen die Befunde die Hauptthese nicht. Sie verbieten jedoch einen breiten Anspruch, erstmals Herkunft und Korrekturgüte gemeinsam zu untersuchen. Matsui et al. (2021) erschließen reale Begutachtungs- und Revisionsverläufe beobachtend; ARIES liefert zugeordnete Kommentare und Änderungen, aber keine experimentelle Prüfung des

vorgeschlagenen Zusammenhangs.

Originalitätskern (was ist neu)

Als möglicher Eigenbeitrag bleibt ein eng abgegrenzter Test: Der Vorteil später Autoridentifikation bei rekonstruierbarer Revisionsgeschichte soll positiv und größer als ohne diese Beziehungen sein, während falsche Änderungen begrenzt bleiben. Eine exakt entsprechende Studie wurde in der heutigen Recherche nicht gefunden. Das ist ein begrenzter Suchbefund, keine Prioritätsgarantie. Insbesondere könnten andere Begriffe wie source disclosure, information lineage, version history oder provenance vergleichbare Arbeiten verdecken.

Die drei numerischen Grenzen machen den Entwurf angreifbar, erzeugen aber für sich keine Originalität. Ihre praktische Begründung fehlt noch. Überträgt man bekannte Transparenzdimensionen auf eine neue Kombination, entsteht zunächst ein Forschungsdesign, keine neue allgemeine Theorie kollektiver Erkenntnis. Die externe Bewertung senkt deshalb Originalität von 6 auf 5. Die Frage bleibt untersuchbar und möglicherweise nützlich, auch wenn ihr begriffliches Grundgerüst bereits ausgearbeitet ist.

Die Suche wurde nach der Synthese und nach der Empirie-Brücke am 13.10.2026 fortgesetzt. Gesucht wurde nach Kombinationen von delayed author identity, revision history, provenance, anonymity und error correction. Thematisch unpassende Treffer zu Datenschutzrichtlinien, DNA-Barcodes und Softwareversionierung wurden nicht als Evidenz übernommen. Das ist eine dokumentierte gezielte Recherche mit selektiver Primärquellenlektüre, kein systematisches Review.

Quellenliste

- Stuart, H. C.; Dabbish, L.; Kiesler, S.; Kinnaird, P.; Kang, R. (2012). [Social transparency in networked information exchange: A theoretical framework](#). CSCW 2012, 451–460. DOI: 10.1145/2145204.2145275.
- Shah, N. B. (2023). [The role of author identities in peer review](#). PLOS ONE 18(6), e0286206. DOI: 10.1371/journal.pone.0286206.
- Rastogi, C.; Song, X.; Jin, Z.; Stelmakh, I.; Daumé III, H.; Zhang, K.; Shah, N. B. (2024). [A Randomized Controlled Trial on Anonymizing Reviewers to Each Other in Peer Review Discussions](#). arXiv, Studie bei UAI 2022. DOI: 10.48550/arXiv.2403.01015.
- Singh, N.; Sovrano, F.; Hellendoorn, V.; Bacchelli, A. (2026). [The Effect of Complexity and Provenance on Code Review Decisions: Evidence from a Controlled Experiment](#). Proceedings of the ACM on Software Engineering 3(FSE), Article FSE158. DOI: 10.1145/3808165.
- Matsui, A.; Chen, E.; Wang, Y.; Ferrara, E. (2021). [The impact of peer review on the contribution potential of scientific papers](#). PeerJ 9, e11999. DOI: 10.7717/peerj.11999.
- D'Arcy, M.; Ross, A.; Bransom, E.; Kuehl, B.; Bragg, J.; Hope, T.; Downey, D. (2024). [ARIES: A Corpus of Scientific Paper Edits Made in Response to Peer Reviews](#). ACL 2024, 6985–7001. DOI: 10.18653/v1/2024.acl-long.377.
- Godlee, F.; Gale, C. R.; Martyn, C. N. (1998). [Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial](#). JAMA 280(3), 237–240. DOI: 10.1001/jama.280.3.237.
- Stelmakh, I.; Shah, N. B.; Singh, A. (2019). [On Testing for Biases in Peer Review](#). Advances in Neural Information Processing Systems 32.

Zugriffsumfang und Aussagegrenzen aller Quellen stehen im getrennten Quellenstudium der Hauptlinie.

Stage 2 — Falsifikationsversuch

Falsifikations-Audit

Die empirische These verlangt drei gleichzeitig erfüllte quantitative Bedingungen. Ein günstiger Einzelvergleich genügt nicht. Ihr Scheitern betrifft diese Fassung im registrierten Anwendungsbereich; die begriffliche Unterscheidbarkeit sachlicher Revisionsbeziehungen und persönlicher Zuschreibungen wäre damit nicht widerlegt. Der primäre Prüfbefund betrifft zunächst die fest gewichtete Aufgabenmischung.

Der Endpunkt bleibt zeitlich unklar: Das gesperrte Ersturteil wird als Messzeitpunkt bezeichnet, die späte Offenlegung folgt danach. Für die behauptete Korrekturwirkung muss ein abschließendes Urteil nach Offenlegung mit einem zwischen den Armen vergleichbaren, festgelegten Revisionsfenster definiert werden. Wird ausschließlich das Ersturteil ausgewertet, erreicht die späte Offenlegung den gemessenen Entscheidungsprozess gar nicht.

Die Nenner bleiben sämtliche zugewiesenen fehlerhaften beziehungsweise richtigen Ausgangsbefunde. Diese sind vom eigenen Ersturteil der Gruppe zu unterscheiden. Eine Beschränkung auf Aufgaben mit falschem Gruppen-Ersturteil würde den Zielparameter wechseln und nach einem möglicherweise bereits durch die Versuchsbedingung veränderten Merkmal auswählen. Auch unveränderte Nenner bestimmen noch nicht, wie fehlende Endurteile bewertet werden; diese Regel muss vorab konkretisiert sein.

Die erlaubten Unterschiede im Erschließungsaufwand lassen eine Erklärung durch unterschiedliche Belastung offen. Eine abgesicherte Interaktion auf der Risikodifferenzskala identifiziert weder einen Statusmechanismus noch gekoppelte Inferenzprozesse.

Konkrete Falsifikationskandidaten

Die folgenden Muster sind mögliche Gegenbefunde, keine berichteten Ergebnisse. L und U bezeichnen die unteren und oberen Grenzen der simultanen 95-Prozent-Konfidenzintervalle.

1. **Unzureichender Verzögerungsnutzen:** Frühe Autorinformation könnte bereits beim Ersturteil helfen; bei rekonstruierbarer Geschichte bleibt der zusätzliche Nutzen später Offenlegung unter der Mindestgröße. Der klare Scheiterpunkt lautet $U_B < 0,05$. Auch ein positiver, statistisch von null unterscheidbarer Effekt rettet diese quantitative Fassung dann nicht.
2. **Unzureichende Interaktion:** Verzögerung hilft mit und ohne verknüpfte Geschichte ähnlich stark. Bei $U_I < 0,03$ scheitert die These selbst dann, wenn B seine Mindestschwelle nachweislich erreicht. Allgemeine Vorteile später Namensnennung ersetzen den vorhergesagten Unterschied zwischen den Effekten nicht. Insbesondere genügt eine lediglich positive Interaktion nicht.
3. **Überschrittene Fehlkorrekturtoleranz:** Späte Offenlegung erhöht bei vorhandener Geschichte die Rate falscher Änderungen so deutlich, dass $L_F > 0,02$ gilt. Damit scheitert die Konjunktion auch bei ausreichenden Werten für B und I. Zusätzlicher Nutzen darf dieses eigenständige Kriterium nicht nachträglich aufwiegen.

Schwellentest

Vorläufiges Bestehen verlangt gleichzeitig $L_B \geq 0,05$, $L_I \geq 0,03$ und $U_F \leq 0,02$. Erfüllt kein Intervall ein Scheiterkriterium, reicht das allein nicht zum Bestehen. Ohne erfüllte Bestehens- oder Scheiterregel bleibt der Versuch unentschieden. Die statistische Entscheidung hängt von tragfähigen Hilfsannahmen über Lösungsschlüssel, Manipulation und Intervallmodell ab; sie liefert keine logische Gewissheit.

Eine entscheidende Unsicherheit betrifft die erreichbare Präzision. Bei wahren Effekten genau an den Grenzwerten lässt sich eine hohe Bestehenswahrscheinlichkeit nicht durch bloße Vergrößerung der Stichprobe erreichen: Die entsprechenden Konfidenzgrenzen müssen vollständig auf der geforderten

Seite liegen. Die Planung muss deshalb plausible Abstände zu den Schwellen berücksichtigen, ohne die Schwellen selbst zu verschieben.

Materialschwellen und zulässige Identitätsleckage sind noch offen. Ihre Festlegung muss vor Ergebniseinsicht erfolgen. Ein nach diesen Kriterien festgestelltes Manipulationsversagen erlaubt keinen Sachentscheid. Ungünstige Effekte nachträglich selbst zum Beleg einer misslungenen Manipulation zu erklären, wäre dagegen Immunsierung; ebenso ergebnisabhängige Ausschlüsse, freie Umklassifizierungen außerhalb des vorgesehenen Einspruchsverfahrens oder weichere Schwellen.

Die praktische Rechtfertigung der Grenzen bleibt ebenfalls offen: Fünf zusätzliche richtige Korrekturen je hundert fehlerhafter Aufgaben und zwei zusätzliche Fehländerungen je hundert richtiger Aufgaben haben unterschiedliche Nenner. Ohne Fehlerhäufigkeiten und Schadensgewichte ergibt sich daraus keine Nutzenbilanz. Diese Begründungslücke beseitigt die Falsifizierbarkeit nicht, begrenzt aber die praktische Bedeutung eines Bestehens.

Stage 3 — Schul-fremde Begutachtung (Ilan Hacking)

Die Hypothese behandelt Revision als Korrektur einer vorab bestimmten Behauptung anhand eines anfechtbaren, aber hinreichend stabilen Lösungsschlüssels. Die Berufung auf Kant, Wittgenstein und Popper rückt explizite Gründe, erkennbare Geltungsstände und Anfechtbarkeit ins Zentrum. Diese traditionsgebundene Auswahl begründet jedoch weder die konkrete Operationalisierung noch die erwartete Effektrichtung. Für die eingegrenzten Methodenaufgaben ist sie vertretbar; die historische Veränderung von Problemen, Beweisstandards und Dokumentationspraktiken bleibt dabei allerdings außerhalb des Versuchs. Eine konkrete Schwäche liegt in der noch offenen Bedeutung des Korrekturendpunkts. Es muss eindeutig sein, ob ein fehlerhafter Aufgabenbefund berichtigt oder ein eigenes fixiertes Ersturteil geändert wird. Frühe Identifikation kann bereits dieses Ersturteil beeinflussen; mehr spätere Korrekturen bedeuten daher nicht zwingend höhere abschließende Richtigkeit. Die Auswertung aller zugewiesenen Gruppen nach richtigen und fehlerhaften Ausgangsbefunden verhindert diese begriffliche Unklarheit allein noch nicht. Außerdem dürfen sich die Geschichtsbedingungen ausdrücklich im Erschließungsaufwand unterscheiden. Bei gleichem Zeitbudget könnte der Zusatznutzen später Namensnennung deshalb davon abhängen, wie viel Prüfzeit die jeweilige Darstellung tatsächlich übrig lässt. Das wäre ein relevanter Verfahrensbefund, erlaubte aber keine eindeutige Zuschreibung an das Verständnis von Revisionsbeziehungen. Die bereits benannte Mechanismusgrenze sollte auch diese Unterschiede der Arbeitspraxis umfassen. Zudem ist Autoridentität kein einheitliches Informationsangebot: Ein unbekannter Name, ein vertrauter Kollege und eine umstrittene Autorität können unterschiedliche Erwartungen auslösen. Gleiche Expertiseangaben beseitigen solche Unterschiede nicht; die Materialprüfungen müssten deren Bedeutung für den registrierten Anwendungsbereich klären. Die gesetzten Nutzen- und Schadensgrenzen betreffen zudem unterschiedliche Ausgangsbefunde. Selbst erfüllte Schwellen garantieren bei überwiegend richtigen Ausgangsbefunden keine höhere gesamte Endgenauigkeit. Ihre angekündigte praktische Begründung muss deshalb das Verhältnis fehlerhafter zu richtigen Befunden und die Folgen falscher Änderungen in einer konkreten Anwendung berücksichtigen. Eine produktive Anschlussidee wäre, die Herstellung des „verlässlichen Prüfers“ selbst zu untersuchen. Historisch ließe sich rekonstruieren, welche Dokumentationspraktiken Änderungen vergleichbar und bestimmten Urhebern zurechenbar gemacht haben. In einer gesonderten Untersuchung könnte anschließend beobachtet werden, was geschieht, wenn Beteiligte ihre Korrekturraten kennen und anhand dieser Werte als zuverlässig eingestuft werden. Sie könnten vorsichtiger Ersturteile abgeben, Fehler anders dokumentieren oder sichtbare Selbstkorrekturen gezielt produzieren. Solche Rückwirkungen wären überprüfbare Möglichkeiten, keine durch den Begriff looping effects bereits gesicherten Tatsachen. Entscheidend wäre, ob die Klassifikation diejenigen Praktiken verändert, deren Qualität sie erfassen soll. Auch eine namenlose Revisionsgeschichte könnte dann zum Träger einer erlernten Reputation werden. Das würde die Strukturthese praktisch zuspitzen, ohne aus dem Leistungstest einen Nachweis ihres Mechanismus zu machen. Aus dem vorgelegten Entwurf folgt weder Bestätigung noch Widerlegung der quantitativen Fassung.

Umgang mit der externen Kritik

Der Audit bestätigt die drei notwendigen Entscheidungsbedingungen und warnt zu Recht vor wechselnden Nennern, nachträglichen Manipulationsausreden und ergebnisabhängigen Schwellen. Seine Forderung, die primäre Korrektur erst nach der späten Namensoffenlegung zu messen, wird dagegen nicht übernommen: Das würde den Zielparameter wechseln. Gemeint ist die Korrektur eines vorgelegten Ausgangsbefunds durch das verbindliche gemeinsame Ersturteil; die späte Gruppe urteilt zu diesem Zeitpunkt bei zurückgehaltener Identität. Die spätere Offenlegung verursacht nicht rückwirkend die gemessene Entscheidung. Die finale Formulierung bezeichnet deshalb ausdrücklich das Zurückhalten der Identität. Ein anschließendes Revisionsfenster wäre ein eigener Folgeversuch. Es wird nicht auf anfänglich falsche eigene Gruppenurteile konditioniert. Übernommen werden eine konkrete Defaultregel für nicht eingereichte Änderungen, Sensitivitätsgrenzen bei technisch unbestimmbaren Endzuständen sowie die Präzisionsplanung mit Abständen zu den beanspruchten Grenzwerten. Alle Änderungen sind Klarstellungen; die drei numerischen Schwellen bleiben unverändert.

Die schul-fremde Prüfung bestätigt die notwendige Trennung von vorgelegtem Befund und eigenem Gruppenurteil; sie wird ausdrücklich festgeschrieben. Darüber hinaus umfasst der mögliche Verfahrenseffekt auch unterschiedlich verbleibende Prüfzeit und Erschließungsaufwand, nicht ausschließlich ein besseres Verständnis von Revisionsbeziehungen. Bekannte, unbekannte und umstrittene Autoren dürfen nicht als austauschbare Identitätshinweise gelten: Die Profile und ihre Vertrautheit gehören zur registrierten Material- und Zielgruppendefinition. Ein Bestehen der drei Schwellen garantiert keine bessere gesamte Endgenauigkeit oder positive Nutzenbilanz; dafür sind die Fehlerhäufigkeit der konkreten Anwendung und die Kosten falscher Änderungen entscheidend. Der Test hält Aufgaben und Richtigkeitsstandards fest und untersucht deshalb keine historische Veränderung solcher Standards. Als eigene spätere Untersuchung bleibt offen, ob sichtbar gemachte Korrekturraten und die Einstufung als verlässlicher Prüfer das Urteil und die Dokumentation selbst verändern. Solche Rückwirkungen werden als prüfbare Möglichkeiten geführt, nicht mit dem Begriff looping effects schon bewiesen.

Bewertung nach externer Prüfung

Kriterium	Intern	Nach externer Prüfung	Begründung
Originalität	6	5	Stuart et al. 2012 nehmen die begriffliche Trennung vorweg; als möglicher Eigenbeitrag bleibt der enge, noch ungeprüfte Interaktionsvergleich.
Falsifizierbarkeit	8	8	Der Messzeitpunkt ist als Wirkung zurückgehaltener Identität geklärt; drei Scheitergrenzen bleiben bestehen, Material- und Präzisionsplanung fehlen.
Begriffliche Klarheit	8	8	Vorgelegter Ausgangsbefund und eigenes Gruppenurteil sind getrennt; institutionell festgelegte Richtigkeitskategorien begrenzen die Übertragung.
Tiefe	7	7	Das Verhältnis von Gründen, Vertrauen und Autorität wird sichtbar, während der Versuch zunächst nur einen begrenzten Leistungseffekt prüft.
Forschungsrelevanz	8	8	Die These schließt an experimentelle Forschung zu Autoridentität, Begutachtung und Fehlererkennung an.
Interdisziplinäre Anschlussfähigkeit	7	7	Metawissenschaft, Sozialpsychologie und Informationsgestaltung haben konkrete Anschlüsse, deren Reichweite durch die eingegrenzten Aufgaben begrenzt bleibt.
Vault-Anschluss	5	5	Vertrauen und die vorige Provenienzfrage liefern Anschlussstellen, aber keine ausgearbeitete Herleitung des behaupteten Effekts.
Antinomie-Test	7	7	Die plausible Gegenthese eines Vorteils früher Verantwortungszurechnung bleibt am selben Korrekturergebnis prüfbar.
Publikationsmöglichkeit	7	7	Ein empirischer Beitrag wäre nach Pilot und registriertem Versuch prüfbar; die vorliegende theoretische Konkretisierung besitzt keinen gesicherten Publikationserfolg.
Summe, gewichtet auf 90	62	61	Keine Wahrheitswahrscheinlichkeit.

Die Hauptthese bleibt empirisch unbestätigt. Der begriffliche Vorläufer begrenzt ihren Neuheitsanspruch; der quantitative Versuch bleibt als enges Forschungsvorhaben offen.

Quellenstudium der Hauptlinie

Vierzehn Quellenhinweise zur namenlosen Fehlerspur. Die Liste umfasst experimentelle und beobachtende Forschung, eine theoretische Arbeit, Methoden- und Korpusarbeiten sowie eine Produktdokumentation. Sie ist kein Bestand von vierzehn Wirksamkeitsnachweisen. Recherche bis 13.10.2026. Die Quellen der Nebenlinie sind im eigenen Abschnitt dokumentiert.

Q1 — Huber, J.; Inoua, S.; Kerschbamer, R.; König-Kersting, C.; Palan, S.; Smith, V. L. (2022)

Nobel and novice: Author prominence affects peer review. PNAS 119(41), e2205779119. DOI: 10.1073/pnas.2205779119.

Anschluss. Dasselbe Manuskript wurde unter unterschiedlichen Zuschreibungen begutachtet; die Ergebnisse zeigen einen Statusbias.

Grenze. Weder richtige Fehlerkorrektur noch Interaktion mit rekonstruierbarer Revisionsgeschichte geprüft.

Tatsächlicher Zugriff. Indexierte Originalpassagen der Verlagsseite zu Abstract, Ergebnissen und Diskussion geprüft; direkter DOI-Aufruf später blockiert. Der zuvor gefundene Chapman-PDF-Link führte am 13.10. auf eine Fehlerseite und wird nicht als funktionierender Volltextzugang ausgewiesen.

Q2 — Tomkins, A.; Zhang, M.; Heavlin, W. D. (2017)

[Reviewer bias in single- versus double-blind peer review.](#) PNAS 114(48), 12708–12713. DOI: 10.1073/pnas.1707323114.

Anschluss. Feldexperiment zur Begutachtung bei WSDM mit Prestigeeffekten.

Grenze. Die methodische Kritik von Stelmakh et al. ist mitzulesen; keine Gleichsetzung von Bewertung und Wahrheit.

Tatsächlicher Zugriff. Originalabstract, Ergebniszusammenfassung und Modellbeschreibung geprüft.

Q3 — Stelmakh, I.; Shah, N. B.; Singh, A. (2019)

[On Testing for Biases in Peer Review.](#) Advances in Neural Information Processing Systems 32.

Anschluss. Ungeeignete Modelle, Messungen und Zuweisungen können Biastests verzerren; die Arbeit entwickelt kontrollierte Testverfahren.

Grenze. Methodische Arbeit ist kein direkter Datentest unserer Gruppenintervention.

Tatsächlicher Zugriff. Autorenabstract und Metadaten geprüft; kein vollständiger Beweismachvollzug.

Q4 — Shah, N. B. (2023)

[The role of author identities in peer review.](#) PLOS ONE 18(6), e0286206. DOI: 10.1371/journal.pone.0286206.

Anschluss. ITCS 2023 verdeckte die Autorennamen zunächst und legte sie nach dem ersten Gutachten offen; spätere Änderungen waren möglich.

Grenze. Kein direkter Richtigkeitstest und kein faktorieller Vergleich rekonstruierbarer Befundgeschichte. Begutachtungsdaten nicht öffentlich; Befragungsdaten verfügbar.

Tatsächlicher Zugriff. Volltext zugänglich; Ablauf, Ergebnisse, Grenzen und Datenverfügbarkeit geprüft.

Q5 — Liu, R.; Jecmen, S.; Conitzer, V.; Fang, F.; Shah, N. B. (2024)

[Testing for reviewer anchoring in peer review: A randomized controlled trial.](#) PLOS ONE 19(11), e0301111. DOI: 10.1371/journal.pone.0301111.

Anschluss. Der untersuchte randomisierte Versuch mit 108 Forschenden ergab keinen signifikanten Ankereffekt.

Grenze. Kein Nachweis allgemeiner Effektfreiheit; die Studie variiert weder unsere Gruppenentscheidung noch beide Faktoren.

Tatsächlicher Zugriff. Originalabstract, experimenteller Ablauf und Ergebnisse geprüft. Zugehöriges öffentliches Repository und README gelesen: Material, anonymisierte Wertungen und Analysecode vorhanden; Daten nicht neu ausgewertet.

Q6 — Godlee, F.; Gale, C. R.; Martyn, C. N. (1998)

Effect on the quality of peer review of blinding reviewers and asking them to sign their reports: a randomized controlled trial. JAMA 280(3), 237–240. DOI: 10.1001/jama.280.3.237.

Anschluss. Bei eingebauten Manuskriptfehlern ergab sich kein signifikanter Unterschied der Fehlererkennung zwischen den untersuchten Begutachtungsverfahren.

Grenze. Andere Intervention und Aufgabensituation; dennoch ein Gegenbefund gegen pauschale Qualitätsversprechen durch Verblindung.

Tatsächlicher Zugriff. Originalabstract sowie Methoden-, Ergebnis- und Diskussionspassagen auf der JAMA-Verlagsseite am 13.10. geprüft.

Q7 — Leek, J. T.; Taub, M. A.; Pineda, F. J. (2011)

Cooperation between Referees and Authors Increases Peer Review Accuracy. PLOS ONE 6(11), e26895. DOI: 10.1371/journal.pone.0026895.

Anschluss. Ein experimentelles Peer-Review-Spiel untersucht, wie Kooperationsanreize die Genauigkeit beeinflussen.

Grenze. Spielmodell; Identität, Kommunikation und Anreize sind nicht automatisch getrennte Ursachen.

Tatsächlicher Zugriff. Originalartikel und ausgewählte Modell- und Ergebnispassagen geprüft.

Q8 — Pleskac, T. J.; Kyung, E. J.; Chapman, G. B.; Urminsky, O. (2025 online; Heft 2026)

Blinded versus Unblinded Review: A Field Study on the Equity of Peer-Review Processes. Management Science 72(4), 2702–2719. DOI: 10.1287/mnsc.2022.01646.

Anschluss. Feldstudie zu 530 Konferenzbeiträgen mit Bewertungsstreuung und Unterschieden zwischen verblindeter und offener Begutachtung.

Grenze. Fairness und Übereinstimmung sind nicht identisch mit der sachlichen Richtigkeit einer Korrektur.

Tatsächlicher Zugriff. Verlagsabstract, Publikationsdaten und Datenverfügbarkeit geprüft; keine vollständige Lektüre behauptet.

Q9 — Matsui, A.; Chen, E.; Wang, Y.; Ferrara, E. (2021)

The impact of peer review on the contribution potential of scientific papers. PeerJ 9, e11999. DOI: 10.7717/peerj.11999.

Anschluss. Beobachtungsanalyse angenommener PeerJ-Arbeiten verbindet Begutachtungsinhalte, Revisionsverläufe und sichtbare Gutachternamen.

Grenze. Selbst gewählte Namensoffenlegung, Selektionsbeschränkung auf angenommene Artikel und andere Zielgrößen verhindern den direkten kausalen Test der Arbeitshypothese.

Tatsächlicher Zugriff. Originalabstract, bibliografische Daten und indexierte Originalpassagen zu Revisionen und Grenzen geprüft. PeerJ-Direktaufruf 403; PMC-Browserprüfung.

Q10 — Elsevier Digital Commons (undatiert; abgerufen 13.10.2026)

Peer Review Process for Journals. Offizielle Produktdokumentation.

Anschluss. Die Plattform unterstützt Revisionsverläufe und das Verbergen identifizierender Dateien bei doppelt verblindeter Begutachtung.

Grenze. Technische Dokumentation, keine Studie zu Wirksamkeit oder gemeinsamer Fehlerkorrektur.

Tatsächlicher Zugriff. Indexierter Originalabschnitt über Revision History und Verbergen identifizierender Dateien gelesen.

Q11 — Rastogi, C.; Song, X.; Jin, Z.; Stelmakh, I.; Daumé III, H.; Zhang, K.; Shah, N. B. (2024)

A Randomized Controlled Trial on Anonymizing Reviewers to Each Other in Peer Review Discussions. arXiv, Studie bei UAI 2022. DOI: 10.48550/arXiv.2403.01015.

Anschluss. Anonymität unter Gutachtern verändert die Nähe abschließender Entscheidungen zu den anfänglichen Bewertungen ranghöherer Gutachter.

Grenze. Identität der Gutachter statt der Autoren; Entscheidungsnähe ist kein objektiver Richtigkeitsmaßstab.

Tatsächlicher Zugriff. Originalabstract und Volltextpassagen zu Zuweisung, Ausfällen und Senioritätseinfluss gelesen.

Q12 — D’Arcy, M.; Ross, A.; Bransom, E.; Kuehl, B.; Bragg, J.; Hope, T.; Downey, D. (2024)

ARIES: A Corpus of Scientific Paper Edits Made in Response to Peer Reviews. ACL 2024, 6985–7001. DOI: 10.18653/v1/2024.acl-long.377.

Anschluss. Verknüpft Gutachterkommentare mit darauf bezogenen Textänderungen.

Grenze. Keine randomisierte Identitätsintervention und kein Goldstandard für Wahrheit jeder Änderung.

Tatsächlicher Zugriff. Originalabstract und bibliografische Angaben der ACL gelesen; Datensatz nicht heruntergeladen oder analysiert.

Q13 — Singh, N.; Sovrano, F.; Hellendoorn, V.; Bacchelli, A. (2026)

The Effect of Complexity and Provenance on Code Review Decisions: Evidence from a Controlled Experiment. Proceedings of the ACM on Software Engineering 3(FSE), Article FSE158. DOI: 10.1145/3808165.

Anschluss. Ein Versuch mit 385 Teilnehmenden variiert Komplexität und Mensch/KI-Herkunftsangabe; berichtet wird kein statistisch gestützter Herkunfts- oder Moderationseffekt.

Grenze. Herkunftsetikettierung ist keine verzögerte persönliche Namensoffenlegung; Komplexität keine Revisionsgeschichte.

Tatsächlicher Zugriff. Originalabstract und Publikationsmetadaten der beteiligten Universität gelesen; kein vollständiger Volltextnachvollzug.

Q14 — Stuart, H. C.; Dabbish, L.; Kiesler, S.; Kinnaird, P.; Kang, R. (2012)

Social transparency in networked information exchange: A theoretical framework. CSCW 2012, 451–460. DOI: 10.1145/2145204.2145275.

Anschluss. Unterscheidet Identitäts- und Inhaltstransparenz und diskutiert Revisionsverläufe, Reputation und deren mögliche gegenläufige Folgen.

Grenze. Theoretischer Rahmen, kein Test der hier spezifizierten numerischen Interaktion.

Tatsächlicher Zugriff. Autoren-PDF: Abstract, Einleitung sowie Abschnitte Identity Transparency und Content Transparency einschließlich Revisionshistorien gelesen. Publikationstitel und DOI zusätzlich auf Sara Kieslers CMU-Publikationsliste verifiziert; PDF trägt abweichenden Untertitel A Framework and Research Question.

Quellenaufnahmen im Vault

Acht selektiv geprüfte Quellennotizen sind für diesen Lauf vorbereitet. Ihre Ablage wird gesondert durch Dateivergleich verifiziert.

- 04 Ressourcen/Philosophen & Werke/Huber 2022 - Autorenstatus im Peer Review
- 04 Ressourcen/Philosophen & Werke/Shah 2023 - Verzögerte Namensoffenlegung im Peer Review
- 04 Ressourcen/Philosophen & Werke/Liu 2024 - Ankerprüfung bei revidierten Gutachten
- 04 Ressourcen/Philosophen & Werke/Godlee 1998 - Verblindung und Fehlererkennung
- 04 Ressourcen/Philosophen & Werke/Matsui 2021 - Begutachtung und Revisionsverläufe
- 04 Ressourcen/Philosophen & Werke/Rastogi 2024 - Anonymität in Gutachterdiskussionen
- 04 Ressourcen/Philosophen & Werke/Singh 2026 - Herkunft und Komplexität beim Code Review
- 04 Ressourcen/Philosophen & Werke/Stuart 2012 - Identitäts- und Inhaltstransparenz

Was aus der Lektüre folgt

Die verzögerte Offenlegung und die Trennung von Identitäts- und Inhaltsgeschichte sind bekannte Ansätze. Offen bleibt der eng definierte gemeinsame Leistungstest. Die vorhandenen Daten liefern Anregungen für Material und Messung; sie ersetzen keinen neuen randomisierten Versuch. Der Blick auf Autorstatus darf nicht mit der Identität anderer Gutachter verwechselt werden. Diese Abgrenzung betrifft unmittelbar die mögliche Deutung eines Effekts.

Reservoir und Rückkopplung

- 06 Hypothesentag/Reservoir/Reservoir - Das Selbst im Verhör 2026-10-12
- 06 Hypothesentag/Reservoir/Reservoir - Überlieferung als Einspruchsrecht 2026-10-12

Rechenweg und Vollständigkeit

Neun Projektkriterien mit ihren tatsächlichen Gewichten: Originalität 1,6; Tiefe und Antinomie-Test je 1,2; übrige Kriterien je 1,0. Die Gewichte summieren sich auf 10,0. Verwendet wird deshalb durchgehend der Normierungsfaktor $9/10 = 0,9$ mit kaufmännischer Rundung. Die veraltete Konfigurationskonstante 10,2 wurde nicht übernommen. Kriteriennamen aus config/kriterien.yaml haben Vorrang vor der beispielhaften Tabelle im Blog-Template. Scores ordnen Kandidaten und sind weder Wahrscheinlichkeiten noch empirische Bestätigungen.

21 vollständige Einzeltexte: sieben Hauptgutachten, sieben Nebenlinien-Kurzgutachten, sieben Repliken nur der Hauptlinie. Dazu fünf Synthesebewegungen, Empirie-Brücke, drei externe Prüfstufen und getrennte Quellenarbeit. Kein OpenRouter-Aufruf.