HYPOTHESENAGENT / WOCHENLAUF

Zeitliche Verknüpfung begünstigt unbegründete Geltungsaberkennung

Vollständiges Gutachten · 2. November 2026

Hauptlinie und Nebenlinie · 7 getrennt erzeugte 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: 68/90 vor dem Panel, 64/90 intern, 61/90 nach externer Prüfung. Empirie-Score: 6/10.

Hypothesentag — 2026-11-02

Vault-Lage und Verfahren

Der Scanner erfasste 520 Dateien, 5 offene Verzweigungen und 68 zeitfensterbezogene Virgin Nodes. Dieser Wochenlauf ist eigenständig auf den 2. November 2026 datiert. Der vorherige unvollständige Lauf vom 30. Oktober bleibt unverändert erhalten. Alle Denkerstimmen sind KI-Rekonstruktionen in getrennten GPT-6-Astra-Kontexten mit maximalem Reasoning. Auch die externe Prüfung bezeichnet getrennte Prüfaufträge und externe Literatur; sie ist keine menschliche Begutachtung oder Prüfung durch einen anderen Modellanbieter.

Verzweigungs-Pickup

Warm-Pick. empirie-chronologie-material aus 06 Hypothesentag/2026-10-19.

Cold-Pick. Schiller - Die Glocke als Regelungsgedicht und hylemorphische Form, Zieldomäne Kulturphilosophie. Virgin bedeutet hier fehlende Verwendung im Suchfenster; eine historische Erstverwendung ist damit nicht nachgewiesen. Der vollständige Seed-Text wurde gelesen und sein begrifflicher Raum erhalten.

Drift und Echo-Chamber. Im Siebentagefenster vom 26. Oktober bis 2. November liegt kein vollständiger Bericht. Es gibt keinen formalen Alarm, aber auch keine Grundlage für einen gemessenen aktuellen Wochentrend. Der ausdrücklich verlangte Außenblick vergleicht daher die letzten vollständigen Wochenläufe vom 5., 12. und 19. Oktober.

Reservoir-Priorität. Autokatalyse als Schwellenarchitektur wurde angeboten, wegen des Vorrangs des Außenblicks nicht konsumiert. Der unerledigte Lauf vom 30. Oktober zählt nicht als neuer publizierter Vorbefund.

Thermalkarte

Fenster 26.10.–02.11.2026: 0 vollständige Berichte. Keine dominante Klassifikation im Zeitfenster; keine gemessene Wochendrift. Kalte Makrothemen: Ästhetik_Goethe, Psychologie_Ethik, Naturwissenschaft_Methode, Anthropologie_Schwelle, Systemtheorie_Luhmann, Erkenntnistheorie_Cassirer. Der vorgeschriebene Außenblick nutzt offen deklariert die drei letzten vollständigen Wochenläufe.

Außenblick — 2. November 2026

Status: verpflichtende Kontrastprüfung der letzten drei vollständigen Wochenläufe; keine gemessene Wochendrift.

Wochenschwerpunkt

Kein gemessener Wochenschwerpunkt: Das Siebentagefenster 26.10.–02.11.2026 enthält keine Berichte. Als ausdrücklich benannter Ersatz dient eine Kontrastprüfung der drei letzten vollständigen Wochenläufe vom 05., 12. und 19.10.2026; diese untersuchen anerkannte Ausstiegsgründe, namenlose Revisionsgeschichte und die Bindung an eigene Arbeit.

Paradigmatische kernannahme

Gemeinsamer Untersuchungszuschnitt, keine nachgewiesene stillschweigende Überzeugung: Ziele, Aufgabenfälle und Erfolgskriterien sind vor der geprüften Entscheidung bestimmt. Untersucht wird vor allem, welche Gründe und Darstellungen die Entscheidung innerhalb dieses Rahmens verändern; die gemeinsame Entstehung neuer Zwecke bleibt außerhalb des Vergleichs.

Gegen these kernsatz

In offenen Kooperationen führt eine kleine, freiwillige und vom Gegenüber tatsächlich nutzbare Vorleistung häufiger zur ausgeführten Zusammenarbeit an einem erst gemeinsam entwickelten Zweck als eine inhaltlich gleiche Absichtserklärung ohne Vorleistung. Der beanspruchte Vorteil besteht auch dann, wenn sich die erwartete Zuverlässigkeit des Gegenübers zwischen den Bedingungen innerhalb vorab begründeter enger Grenzen nicht unterscheidet.

Alternative tradition

Pragmatistische Sozialphilosophie in der Anschlussrichtung Follett und Dewey; zweckbildende Kooperation als Alternative zur ausschließlichen Prüfung bereits bestimmter Optionen.

Aufhaenge knoten

Luhmann - Vertrauen als Komplexitätsreduktion

Vorgeschlagene falsifikation

Die konjunktive These scheitert im registrierten Kooperationsbereich, wenn bei gültiger Durchführung und ausreichender Präzision entweder der praktisch relevante Vorteil der Vorleistung für tatsächlich ausgeführte neue gemeinsame Vorhaben ausgeschlossen wird oder der Erwartungsunterschied die vorab festgelegte Äquivalenzgrenze überschreitet. Ein Kooperationsvorteil allein, ein lediglich nicht signifikanter Unterschied der Erwartungen oder die nachträgliche Auswahl von Paaren mit unveränderter Erwartung bestätigen sie nicht.

Begruendung aussenblick

Der verpflichtende Außenblick erweitert den zuletzt wiederkehrenden Untersuchungsrahmen um eine ernsthafte pragmatistische Alternative. Er ist eine Kontrastprüfung nach drei vollständigen Läufen und keine Drift-Messung im leeren Siebentagefenster; die früheren Hypothesen werden nicht zu einer generellen Transparenz- oder Kontrollthese umgedeutet.

Neuheitsgrenze

Ko-Produktion neuer Ziele durch Vertrauen ist als Vorarbeit mitgeteilt; eigener enger Prüfgegenstand ist tatsächliche Vorleistung gegen Absichtserklärung, mit ausgeführtem neuem gemeinsamen Vorhaben und einem Äquivalenzkriterium für Zuverlässigkeitserwartungen.

Pickup und Reservoir

Der neue Block devil_advocate_pick wurde dem bestehenden Pickup hinzugefügt. Alle vorherigen Felder bleiben unverändert. Nach der verbindlichen Priorität wird H3 aus diesem Außenblick entwickelt; der Reservoir-Kandidat wird nicht konsumiert.

Datengrundlage

Referenzläufe: 2026-10-05, 2026-10-12 und 2026-10-19. Berichte im Siebentagefenster 2026-10-26 bis 2026-11-02: 0. Der Außenblick behauptet keine Verteilung oder Häufigkeit für diese leere Woche und keine Widerlegung der engeren Vorlaufthesen.

Initiale Hypothesen — 2. November 2026

Modell: GPT-6 Astra · Reasoning: max · Status: Generierung, noch ohne unabhängige Prüfung oder Bewertung.

Der Außenblick vergleicht die vollständigen Wochenläufe vom 5., 12. und 19. Oktober. Das aktuelle Siebentagefenster ist leer; eine gemessene Wochendrift wird nicht behauptet. Das Reservoir bleibt ungenutzt.

H1 — Die Chronologie erfindet einen Widerruf

Kernsatz. Bei gleichzeitig sichtbaren, sachlich unabhängigen und weiterhin gültigen Belegen zu einem zeitlich unveränderten Sachverhalt erhöht ihre Verknüpfung nach bloßen Erfassungszeiten gegenüber einer unverknüpften Darstellung die Häufigkeit unbegründeter Ersetzungsurteile: Später erfasste Belege werden als Aufhebung früherer behandelt. Welche Belege dadurch Geltung verlieren, folgt dem vertauschten Zeitetikett, obwohl der sachlich richtige Umgang mit ihnen unverändert bleibt.

Begründung. Die Verzweigung empirie-chronologie-material aus 06 Hypothesentag/2026-10-19 verlangt eine Darstellung zeitlicher Herkunft ohne zusätzliche Schlussinformation. Die neue These trennt die im Material ausdrücklich enthaltene Information von einer erst durch seine Darstellung erzeugten Zuschreibung: Eine rein zeitliche Verbindung könnte als sachliche Revision gelesen werden. Anders als die Besitzthese vom 19. Oktober prüft sie weder eigene gegen fremde Arbeit noch einen Vorzeichenwechsel der Abschlussrichtigkeit, sondern eine bestimmte Fehlklassifikation des Geltungsstands. Die im Eingangsmaterial beschriebene Schwierigkeit wirklicher Rücknahmen motiviert diese Gegenfrage, belegt aber keinen Effekt fingierter Rücknahmen.

Falsifikationsbedingung. Die These scheitert im vorab festgelegten Aufgabenbereich, wenn bei unabhängig validiertem Material und hinreichend präzisen Vergleichen entweder der relevante zusätzliche Anteil falscher Ersetzungsurteile durch die zeitliche Verknüpfung ausgeschlossen wird oder sich die Zuordnung des vermeintlich aufgehobenen Belegs beim Vertauschen der Zeitetiketten nicht entsprechend mitverschiebt. Ein bloßer Unterschied der Endantworten, breite Intervalle oder das nachträgliche Auswählen nur der Personen mit missverstandenem Geltungsstand entscheiden diesen Test nicht.

Quelle. warm_pick: empirie-chronologie-material — 06 Hypothesentag/2026-10-19; bereitgestellter Empirieabschnitt dieses Berichts.

Thematischer Raum. Epistemologie und Urteilspsychologie: Darstellungsform, zeitliche Herkunft und Geltungsstatus von Evidenz.

Qualitätsvermerk. Eigenständiger Prüfgegenstand gegenüber dem 19.10.: unbegründete Ersetzungsrelation statt Eigentumswirkung. Die Einordnung als Effekt der Chronologie ist zu prüfen; eine allgemeine Neuheitsbehauptung gegenüber Forschung zu Reihenfolge und Aktualität wird nicht erhoben. Noch fehlen ein validiertes Material, eine praktisch begründete Mindestwirkung und eine Präzisionsplanung.

Konkretisierung des Prüfwegs

• **Material:** Kleine, nach vorab geprüften Regeln auswertbare Evidenzaufgaben. Die Belege betreffen denselben unveränderten Zustand, besitzen erkennbare unabhängige Quellen und sind sämtlich weiterhin gültig. Die Erfassungszeiten bezeichnen ausschließlich eine sachlich irrelevante Ablagefolge, keine Messzeit, Qualitätsstufe oder Revision.

• **Vergleich:** Randomisierte Verknüpfung versus unverknüpfte Übersicht; in beiden Darstellungen stehen dieselben Belege, Zeitetiketten und ausdrücklichen Geltungsangaben gleichzeitig bereit. Die Zuordnung früher und später Erfassungszeiten zu den Belegen wird unabhängig variiert. Seitenposition, Lesbarkeit und technische Auffindbarkeit werden unabhängig vorgeprüft.

• **Primaerer endpunkt:** Anteil vorab definierter falscher Entscheidungen, einen gültigen Beleg wegen eines anderen Belegs als aufgehoben zu behandeln. Die gleiche Statusfrage und Auswertung gelten für alle zugewiesenen Personen und Aufgabenfälle.

• **Grenze:** Ein solcher Befund wäre ein Darstellungsbefund. Er bewiese ohne zusätzliche Untersuchung weder einen Eigentumsmechanismus noch eine allgemeine Irrationalität von Chronologie; tatsächliche Rücknahmen und zeitlich veränderliche Sachverhalte liegen außerhalb des beanspruchten Bereichs.

H2 — Wer die Form zerbrechen darf

Kernsatz. Der kulturphilosophische Gehalt von Schillers Lied von der Glocke liegt in einer asymmetrischen Autorität zum Formwechsel: Der sachkundige Meister darf die Form zur rechten Zeit zerstören, während der unkontrollierte Ausbruch aus ihr als Zerstörung gemeinsamer Ordnung erscheint. Der Stil des Gedichts überträgt damit ein handwerkliches Urteil über den Zeitpunkt auf die Frage, wer zur gesellschaftlichen Veränderung berechtigt ist.

Begründung. Schiller - Die Glocke als Regelungsgedicht und hylemorphische Form stellt die Formung von Erz und bürgerlichem Leben nebeneinander und begrenzt ausdrücklich die Übertragbarkeit dieses poetischen Bildes auf eine allgemeine Theorie. Die vom Hauptlauf geprüften Textpassagen verschärfen diese Grenze: Formzerstörung wird sowohl als notwendiger Arbeitsschritt als auch als bedrohliche Entfesselung dargestellt. Die Hypothese untersucht, ob der Wechsel zwischen Arbeitsbeschreibung und gesellschaftlichem Urteil die Autorität des Meisters über seinen technischen Zuständigkeitsbereich hinaus verlängert. Ihr Gegenstand ist diese poetische Übertragung von Urteilshoheit, nicht ein bereits nachgewiesener Einfluss auf historische Leser oder die politische Gesinnung des Autors.

Falsifikationsbedingung. Die Lesart scheitert, wenn eine quellennahe Analyse der Übergänge zwischen Gussbeschreibung und Gesellschaftspassagen zeigt, dass das Gedicht die handwerkliche Urteilshoheit gerade nicht auf soziale Veränderung überträgt, oder wenn eine konkurrierende Deutung beide Arten des Formbruchs und die Sprecherwechsel mit weniger unbelegten Zusatzannahmen erklärt. Einzelne ähnliche Wörter oder der isolierte Hinweis auf einen konservativen Inhalt genügen nicht als Bestätigung.

Quelle. zieldomaene: Kulturphilosophie; cold_pick: Schiller - Die Glocke als Regelungsgedicht und hylemorphische Form. Zusätzlich vom Hauptlauf bereitgestellter Befund zum Originaltext in Das Lied von der Glocke (1800).

Thematischer Raum. Kulturphilosophie und literarische Hermeneutik: Stil, poetische Autorität und Legitimation gesellschaftlicher Formveränderung.

Qualitätsvermerk. Konkreter Textgegensatz statt Übertragung von Erkenntnisexperimenten auf Literatur. Die Behauptung einer politischen Autoritätsübertragung bleibt eine prüfbare Auslegung; ihr Neuheitsgrad gegenüber der Schiller-Forschung ist offen. Der Pickup-Status virgin_node gilt nur für das Suchfenster und belegt keine historische Erstverwendung.

Konkretisierung des Prüfwegs

• **Primaermaterial:** Schiller: Das Lied von der Glocke, Textfassung des Musen-Almanachs für das Jahr 1800, S. 251–272, vom Hauptlauf über Wikisource geprüft.

• **Url:** https://de.wikisource.org/wiki/Das_Lied_von_der_Glocke_(1800)

• **Pruefschritte:** Die Passagen des absichtlichen Formzerbrechens, des unbeherrschten Ausbruchs sowie die Übergänge von Arbeitsbeschreibung zu gesellschaftlichen Urteilen werden als zusammenhängender Textvergleich rekonstruiert. Geprüft werden Sprecher, angesprochene Handelnde, zeitliche Bedingungen und Begründung der jeweiligen Erlaubnis oder Verurteilung.

• **Konkurrenzlesart:** Eine engere Deutung könnte ausschließlich den Unterschied zwischen sachkundiger und unbeherrschter Handlung beschreiben, ohne daraus eine gesellschaftliche Verteilung von Veränderungsrechten abzuleiten. Diese Alternative muss am selben Textbestand geprüft werden.

• **Grenze:** Keine pauschale Gleichsetzung von Formbruch und Freiheit; keine ungeprüfte Rezeptionsgeschichte, kein behaupteter Kanonisierungsmechanismus.

H3 — Vertrauen stiftet einen gemeinsamen Zweck

Kernsatz. In offenen Kooperationen führt eine kleine, freiwillige und vom Gegenüber tatsächlich nutzbare Vorleistung häufiger zur ausgeführten Zusammenarbeit an einem erst gemeinsam entwickelten Zweck als eine inhaltlich gleiche Absichtserklärung ohne Vorleistung. Der beanspruchte Vorteil besteht auch dann, wenn sich die erwartete Zuverlässigkeit des Gegenübers zwischen den Bedingungen innerhalb vorab begründeter enger Grenzen nicht unterscheidet.

Begründung. Eine pragmatistische Lesart von Vertrauen fragt, welche gemeinsamen Zwecke im Handeln erst entstehen; Luhmann - Vertrauen als Komplexitätsreduktion liefert den konkreten Anschluss über Handlungsfähigkeit unter fortbestehender Unsicherheit. Die drei jüngsten vollständigen Wochenläufe prüfen Entscheidungen an schon bestimmten Zielen, Aufgaben und Geltungsregeln, während diese Gegenthese die Entstehung eines gemeinsamen Zwecks zum Gegenstand macht. Die vom Hauptlauf recherchierte Verbindung von Follett, Dewey, Vertrauen und der gemeinsamen Bildung neuer Ziele ist eine Vorarbeit und ausdrücklich keine Neuheit dieser Hypothese. Ihr enger eigener Prüfgegenstand ist der Vergleich einer vollzogenen Vorleistung mit einer bloßen Absichtserklärung bei gleicher späterer Ausstattung und zugleich ausbleibendem relevanten Unterschied der Zuverlässigkeitserwartung.

Falsifikationsbedingung. Die konjunktive These scheitert im registrierten Kooperationsbereich, wenn bei gültiger Durchführung und ausreichender Präzision entweder der praktisch relevante Vorteil der Vorleistung für tatsächlich ausgeführte neue gemeinsame Vorhaben ausgeschlossen wird oder der Erwartungsunterschied die vorab festgelegte Äquivalenzgrenze überschreitet. Ein Kooperationsvorteil allein, ein lediglich nicht signifikanter Unterschied der Erwartungen oder die nachträgliche Auswahl von Paaren mit unveränderter Erwartung bestätigen sie nicht.

Quelle. devil_advocate — verpflichtende Kontrastprüfung der vollständigen Wochenläufe 05., 12. und 19.10.2026; Vault-Anker Luhmann - Vertrauen als Komplexitätsreduktion.

Thematischer Raum. Pragmatistische Sozialphilosophie und Kooperationsforschung: tätiges Vertrauen und gemeinsame Bildung von Zwecken.

Qualitätsvermerk. Begrenzt den eigenen Beitrag auf einen konkreten Verfahrensvergleich. Vertrauen als produktive Bedingung neuer Ziele und mögliche Kosten von Kontrolle sind bereits literarisch vorgezeichnet. Ein Effekt würde den Beitrag von Gegenseitigkeit, Selbstbindung, Besitzverteilung oder

Signalen noch nicht isolieren; Luhmanns Komplexitätsreduktion wird nicht mit einer bloßen Senkung subjektiver Ausfallwahrscheinlichkeit gleichgesetzt.

Gegenthese zu. Dem gemeinsamen Untersuchungszuschnitt der drei letzten vollständigen Wochenläufe: Handlungsfähigkeit wird dort an bereits bestimmten Alternativen und Richtigkeitsmaßstäben untersucht. Die Gegenthese wechselt zur gemeinsamen Entstehung von Zwecken und widerlegt damit keine der engeren früheren Wirkungsthesen.

Konkretisierung des Prüfwegs

• **Aufgabe:** Paare bearbeiten eine offene Gestaltungsaufgabe. Vorher werden ihre je einzeln bevorzugten Zwecke festgehalten. Ein neues gemeinsames Vorhaben muss von beiden zuvor allein gewählten Zwecken substanziell abweichen, von beiden angenommen und mit einer vorab bestimmten realen Folgehandlung begonnen werden; blind kodierte Neuheit allein reicht nicht.

• **Vergleich:** Die Untersuchungseinheiten werden einer Möglichkeit tatsächlicher freiwilliger Vorleistung oder einer inhaltlich entsprechenden Absichtserklärung zugewiesen. Die Vorleistung soll vor der gemeinsamen Zweckbildung erfolgen; zum Beginn der eigentlichen Zusammenarbeit müssen Ausstattung, verfügbare Handlungsmöglichkeiten und Informationsinhalt vergleichbar sein. Alle zugewiesenen Paare bleiben in der primären Auswertung, auch wenn eine angebotene Vorleistung unterbleibt.

• **Zuverlaessigkeit:** Erwartete Zuverlässigkeit wird vor der Zuweisung und vor dem gemeinsamen Arbeitsbeginn in beiden Bedingungen identisch erhoben. Ein Äquivalenztest zwischen den randomisierten Bedingungen ersetzt die ungeeignete Auswahl einzelner Paare ohne Erwartungsänderung.

• **Offener designpunkt:** Eine glaubwürdige freiwillige Vorleistung lässt sich nicht ohne Weiteres randomisiert erzwingen. Ein unabhängiger Pilot muss deshalb klären, ob die Zuweisung der Gelegenheit zur Vorleistung einen sauberen und hinreichend starken Vergleich ermöglicht, ohne zugleich Ressourcen oder den sachlichen Möglichkeitsraum ungleich zu machen. Solange dies nicht gelingt, liegt ein Forschungsentwurf und kein fertiges Experiment vor.

• **Grenze:** Die These betrifft die Wirkung des zugewiesenen Verfahrens und seine gemeinsame Ergebniskonstellation. Sie erlaubt keine isolierte Zuschreibung an einen inneren Vertrauensmechanismus und keine allgemeine Widerlegung systemtheoretischer Vertrauensbeschreibungen.

Mitgeteilte Literaturgrenzen

• Vom Hauptlauf mitgeteilte Neuheitsgrenze: Heenen/Brown (2026) verbinden Follett, Dewey, Vertrauen und gemeinsame Bildung neuer Ziele beziehungsweise Lösungen. Hier keine Aussage zum vollständigen Artikel über den mitgeteilten Befund hinaus. Quelle

• Vom Hauptlauf geprüfter Abstract: Falk/Kosfeld (2006), The Hidden Costs of Control. Kontrolle als Misstrauenssignal und mögliche Leistungseinbußen sind Vorarbeiten, kein Beleg des vorliegenden Vergleichs. Quelle

Diversität und Status

• **H1:** Konkrete Fehlzuschreibung sachlicher Geltung durch zeitliche Darstellung; eigener Prüfgegenstand gegenüber Eigentums- und Erwerbsvergleich vom 19.10.

• **H2:** Kulturphilosophische Textauslegung der Autorität zum Formwechsel mit Stil als explizitem Ankerbegriff. Der Gedichttext und seine Sprecher- und Handlungskonstellation sind nicht durch

Evidenzaufgaben ersetzbar.

• **H3:** Pragmatistische Kooperationsfrage nach der Bildung neuer gemeinsamer Zwecke; die Aufgabe besitzt gerade kein vorab richtiges Ziel. Tatsächliche Vorleistung und die Ausführung eines gemeinsam entwickelten Vorhabens sind nicht durch zeitliche Belegverknüpfung ersetzbar.

• **Methodische grenze:** H1 und H3 sind empirische Forschungsvorschläge mit verschiedenen Gegenständen; H2 ist eine hermeneutische These. Eine Vielfalt aus drei verschiedenen Methodentypen wird damit nicht behauptet.

• **Literaturgrenze:** Unabhängige Neuheits- und Quellenprüfung bleibt erforderlich. Es werden keine empirischen Effekte, historischen Rezeptionsbefunde oder Erstmaligkeitsnachweise erfunden.

Keine Punktwerte vergeben. Keine Professorrolle übernommen. Der Suchfensterstatus des Schiller-Knotens wird nicht als Nachweis historischer Erstverwendung behandelt.

Stufe 3: Kritischer Professor — 02.11.2026

Unabhängige Prüfung der drei Ausgangshypothesen nach dem Skill kritischer-professor, eine inhaltliche Prüfiteration einschließlich der vor Abschluss nachgereichten H2-Quellenbefunde. Modellangabe gemäß Auftrag: GPT-6 Astra; Reasoning: max.

Die Prüfung verwendet hypothesen-initial-2026-11-02.json, literatur-vorrecherche-2026-11-02.md, literatur-ergaenzung-2026-11-02.md, die ausdrücklich nachgereichten Dateien literatur-schiller-2026-11-02.md/.json und den beauftragten Skill. Es erfolgten keine neuen Quellen-, Vault- oder Fremddienstzugriffe. Die dokumentierten Literaturbefunde werden innerhalb ihrer angegebenen Zugriffstiefe behandelt.

Methodische Korrektur gegenüber einer Formulierung im Skill: Eine ungestützte Hypothese ist vorläufig und unbestätigt, nicht deshalb falsch. Die hermeneutische These H2 wird anhand konkurrierender Textdeutungen geprüft. Es werden keine Bewertungen, Punktwerte oder Rangfolgen vergeben. Die Reformulierungen sind Vorschläge zur gesonderten Übernahme durch den Hauptlauf.

H1 — Die Chronologie erfindet einen Widerruf

Die prüfbare Behauptung ist eine Konjunktion: Grafische Verknüpfung nach sachlich irrelevanten Erfassungszeiten erhöht falsche Ersetzungsurteile gegenüber denselben unverknüpften Belegen; der zusätzliche Fehler richtet sich nach dem früheren Zeitetikett. Die Gegenhypothese H□ lautet, dass mindestens eine dieser beiden Vorhersagen ausbleibt. Ein bloßer Unterschied der Abschlussantworten prüft diese Behauptung nicht. Die Unabhängigkeit und fortdauernde Gültigkeit der Belege müssen durch Aufgabenregeln eindeutig sein; unabhängige Quellen allein garantieren weder gleichen Informationswert noch fehlende Widersprüche.

Für den behaupteten Effekt liegen keine eigenen Daten vor. Die bereitgestellten Vorrecherchen referieren verwandte Befunde zur Präsentationsfolge, zu wechselnden Kausalbeziehungen und zu Rücknahmen. Ecker und Kollegen trennen nach dem mitgeteilten Befund Darbietungsposition und Aktualität; sie untersuchen jedoch echte konkurrierende Erklärungen und Rücknahmen. Daraus folgt weder Bestätigung noch Widerlegung einer fingierten Ersetzungsrelation durch Ablagezeiten. Effektstärke, Ausgangsfehlerrate und Übertragbarkeit sind unbekannt; die begrenzten Quellenzugriffe erlauben keine quantitative Ableitung.

Die Zeitordnung der Labels muss unabhängig von tatsächlicher Erstexposition und letzter Darbietung variiert werden. Gleichzeitige Sichtbarkeit in der Endübersicht beseitigt vorherige Aktualitätseffekte nicht.

Eine gemeinsame kontrollierte Exposition vor der Übersicht und eine unabhängig randomisierte Expositionsreihenfolge machen die Unterscheidung prüfbar. Ein durch die Grafik selbst ausgelöster Blick- oder Leseverlauf bleibt dagegen möglicher Mechanismus des Darstellungsverfahrens; seine nachträgliche statistische Entfernung wäre keine saubere Isolation.

Drei Konkurrenzdeutungen sind zu prüfen: gerichtete Pfeile können unabhängig von Zeit eine Abhängigkeit suggerieren; Aufmerksamkeit oder Lesereihenfolge können einen Beleg bevorzugen; Versuchspersonen können Erfassungszeit trotz Aufgabenregel für Aktualität der Sache halten. Eine geometrisch vergleichbare Verknüpfung ohne zeitliche Bedeutung hilft bei der ersten Abgrenzung; Verständnis wird unabhängig vorgeprüft, nicht als nachträglicher Ausschlussgrund benutzt.

Quantitativ sind absolute Unterschiede der Fehlerraten und der Fehlerzuordnung mit Unsicherheitsintervallen erforderlich. Wiederholte Urteile derselben Personen und Aufgaben dürfen nicht als unabhängig gezählt werden. Mindestwirkungen und Präzisionsplanung benötigen einen separaten Pilot; breite Intervalle entscheiden nichts. Entscheidend fehlt bisher die identifizierende Trennung zwischen bloßem Zeitetikett, wirklicher Darbietung und allgemeiner Verbindungsgrafik. Die Hypothese ist vorläufig, nicht deswegen falsch.

Entscheidende Kritik: Der Entwurf kann einen Gesamtunterschied zwischen Darstellungen finden, ohne den zusätzlichen Fehler einer irrelevanten Erfassungszeit statt tatsächlicher Aktualität oder allgemeiner Pfeilsemantik zuordnen zu können.

Konkrete Abhilfe: Grafische Verknüpfung, Zuordnung der Erfassungslabels und tatsächliche Expositionsreihenfolge unabhängig variieren; eine passende Verbindungskontrolle vorsehen und den Zuwachs an gezielt gegen früher etikettierte Belege gerichteten Ersetzungsfehlern vorab auswerten.

H2 — Wer die Form zerbrechen darf

Die engere hermeneutische These lautet: Die Verbindung von Meisterurteil, Arbeitsauftrag und Aufruhrpassage lässt gesellschaftliche Veränderung als Frage legitimer Leitung erscheinen. Ihre zentrale Konkurrenz erklärt denselben Gegensatz durch sachgerechte Arbeitsfolge und unterschiedliche Materialzustände, ohne eine gesellschaftliche Verteilung von Veränderungsrechten anzunehmen. Eine statistische Nullhypothese wäre hier unpassend; entscheidend sind Textdeckung, Erklärungskraft und notwendige Brückenannahmen.

Die nachgereichte Quellenprüfung korrigiert drei Voraussetzungen: Der Originaltext sagt kann, nicht darf; die Gesellen erhalten den Auftrag zur Zerstörung; der erlaubte Arbeitsschritt am fertigen, erkalteten Werk und der Ausbruch glühenden Erzes betreffen unterschiedliche Zustände. Die Ausgangsfassung macht damit aus technischem Können zu unmittelbar normative Berechtigung und verkürzt die Handlungskonstellation auf den Meister. Die Delegation an Gesellen widerlegt eine Zuständigkeitslesart nicht, verändert sie aber: Zu prüfen wäre die Autorität zur Anweisung, nicht ein exklusives Recht des Meisters auf die ausführende Handlung.

Die bereitgestellte Schiller-Vorrecherche dokumentiert selektiv geprüfte Primärtextstellen. Luserke-Jaqui verbindet bereits Meisterwarten und gesellschaftliche Selbstbesinnung, unterscheidet aber Beschreibung von Erziehungsprogramm. Breu behandelt Formfindung und Meistervorbild; Hahn ist nur über Metadaten und Abstract erschlossen. Grundanalogie und gesellschaftliche Zeitlichkeit sind damit vorgezeichnet. Die enger gefasste Zuständigkeitsfrage ist kein nachgewiesenes Forschungsdesiderat. Diese Prüfung hat keine dieser Quellen zusätzlich selbst geöffnet.

Die technische Zeitfolge ist eine reale Konkurrenz zur normativen Lesart: Erst Vollendung und Abkühlung machen den angeordneten Formbruch sachgerecht. Der Unterschied der Folgen kann also am Werkzustand liegen, nicht an einer bevorzugten Personengruppe. Zwei weitere Alternativen sind eine

Bewertung produktiver gegenüber zerstörerischen Ergebnissen und eine rhetorische Warnung vor Kontrollverlust ohne gesellschaftliche Kompetenzordnung. Alle drei müssen denselben Textbestand einschließlich Gegenstellen erklären. Politische Wertung und benachbarte Strophen allein entscheiden den Vergleich nicht. Sprechhaltungswechsel dürfen zudem nicht ungeprüft als Sprecherwechsel gelten.

Quantitative Effektgrößen sind hier ohne Gegenstand. Nötig ist eine vollständige Übersicht der einschlägigen Modalverben, Imperative, Akteure, Werkzustände und Übergänge; Wortzählung ersetzt keine Interpretation. Die entscheidende Abhilfe ist eine Aussage über textinterne Autorisierung geleiteten Handelns, deren Brücke vom Können zur Zuständigkeit erst belegt werden muss. Historische Leserwirkung und Autorenabsicht bleiben ausgeschlossen. Die Lesart ist vorgeschlagen, weder bestätigt noch durch diese Einwände bereits widerlegt.

Entscheidende Kritik: Die Ausgangsfassung liest kann als darf, übersieht die beauftragten Gesellen und vergleicht verschiedene Werkzustände. Damit ist eine asymmetrische Veränderungsberechtigung noch nicht aus dem Text hergeleitet; Grundanalogie und gesellschaftliche Zeitlichkeit sind zudem bereits literarisch erschlossen.

Konkrete Abhilfe: Als engere Lesart urteilsfähiger Leitung und beauftragten Formbruchs formulieren. Können, Anweisung, Ausführung, Zeitpunkt und Materialzustand getrennt rekonstruieren und gegen sachgerechte Arbeitsfolge, Folgenbewertung und Kontrollverlust prüfen; keine belegte Forschungslücke behaupten.

H3 — Vorleistung eröffnet einen gemeinsamen Zweck

Die Ausgangsfassung vermischt zwei Zielgrößen: die Wirkung tatsächlich freiwillig geleisteter Vorarbeit und die Wirkung eines zufällig angebotenen Vorleistungsverfahrens. Identifizierbar ist zunächst nur Letztere. Präzise lautet H□: Die Zuweisung dieses Verfahrens erhöht gegenüber dem Erklärungsverfahren die Wahrscheinlichkeit eines tatsächlich begonnenen neuen gemeinsamen Vorhabens mindestens im vorab begründeten Umfang; zugleich bleiben die mittleren Zuverlässigkeitserwartungen zwischen den Verfahren innerhalb begründeter Äquivalenzgrenzen. H□ ist die Negation mindestens einer dieser Bedingungen. Ausbleibende Signifikanz eines Erwartungsunterschieds genügt nicht.

Die bereitgestellten Quellen liefern keinen Test dieser Konjunktion. Heenen und Brown behandeln nach der dokumentierten selektiven Lektüre bereits die gemeinsame Bildung neuer Ziele durch Vertrauen. Edmondsons Zusammenhang von psychologischer Sicherheit und Lernen ist kein randomisierter Vorleistungsvergleich. Falk und Kosfeld prüfen Kontrolle; Goeschl und Jarke prüfen Beobachtbarkeit und Inspektionskosten. Die Übertragung auf tatsächlich begonnene, neu entwickelte Zwecke bleibt offen. Eine eigenständige Versuchsanordnung ist möglich, ein allgemeiner Neuheitsanspruch nicht belegt.

Die kausale Reihenfolge verlangt: individuelle Zwecke und Ausgangserwartungen erfassen, Verfahren zufällig zuweisen, mögliche Vorleistung beziehungsweise Erklärung durchführen, Erwartungen erneut messen und erst danach gemeinsam Zwecke entwickeln. Tatsächlich Leistende nachträglich mit Erklärenden zu vergleichen würde Selbstselektion einführen: Kooperationsbereitschaft könnte sowohl Vorleistung als auch gemeinsames Handeln verursachen. Die primäre Analyse muss deshalb alle zugewiesenen Paare einschließen. Vorleistungsquote und Annahme werden als Durchführung dokumentiert; sie ersetzen keinen Nachweis des kausalen Effekts der freiwilligen Handlung.

Drei Alternativen zum Vertrauensmechanismus bleiben: Gegenseitigkeitsverpflichtung, Selbstbindung durch eigenen Aufwand und ein durch frühe Nutzbarkeit oder Inhalte erleichterter gemeinsamer Arbeitsbeginn. Gleiche Ausstattung beim Arbeitsstart beseitigt frühere Zeit-, Aufmerksamkeits- oder Informationsvorteile nicht automatisch. Ressourcen, Inhalte und Kontaktzeit müssen praktisch

abgeglichen werden; verbleibende Unterschiede gehören zum geprüften Verfahren.

Quantitativ fehlen Grundrate, Vorleistungsquote, Mindestwirkung und begründete Äquivalenzgrenzen. Erforderlich sind absolute Unterschiede mit Intervallen auf Paarebene und eine Planung für beide Bestandteile der Konjunktion. Erwartungsäquivalenz der Gruppen schließt Vermittlung über individuelle Erwartungsänderungen, gegenläufige Teilgruppen oder Messfehler nicht aus. Weder ein Intervall, das die Äquivalenzgrenze nur berührt, noch ein unpräziser Kooperationseffekt widerlegt allein die These. Entscheidend ist die Umformulierung zum Verfahrenseffekt; der Zweckbildungsanspruch bleibt bestehen.

Entscheidende Kritik: Randomisierte Gelegenheit identifiziert nicht die Wirkung tatsächlich freiwilliger Vorleistung. Außerdem erlaubt Äquivalenz mittlerer Erwartungswerte keinen Ausschluss einer Vermittlung über Erwartungen.

Konkrete Abhilfe: Den Kernsatz als Effekt der zufälligen Verfahrenszuweisung formulieren, alle Paare nach Zuweisung auswerten und Erwartungsäquivalenz nur als zusätzliche Ergebniskonstellation prüfen. Tatsächliche Vorleistung und mögliche Mechanismen gesondert als offene Erklärungsfragen dokumentieren.

Übergabe

Die Prüfung ist abgeschlossen. Die korrigierten Fassungen stehen in reformulierung-vorschlag-2026-11-02.json; die Ausgangshypothesen bleiben unverändert. H1 bewahrt den erfundenen Geltungswechsel, H2 die Autorisierung von Formveränderung und H3 die gemeinsame Bildung neuer Zwecke. Noch offene Material-, Text- und Pilotprüfungen sind ausdrücklich benannt.

Reformulierung nach dem Kritischen Professor

Die unabhängig erarbeiteten Reparaturen wurden vom Hauptlauf sachlich geprüft und für die anschließende Bewertung übernommen. Die drei Ausgangsfassungen bleiben getrennt dokumentiert. Die Übernahme ist eine redaktionelle Arbeitsentscheidung und keine Bestätigung der Hypothesen.

H1 — Die Chronologie erfindet einen Widerruf

Kernsatz. Bei weiterhin gültigen, sachlich unabhängigen Belegen zu einem unveränderten Sachverhalt erhöht eine grafische Verknüpfung nach sachlich irrelevanten Erfassungszeiten gegenüber einer informationsgleichen unverknüpften Übersicht falsche Ersetzungsurteile. Der zusätzliche Fehler richtet sich gegen den jeweils früher etikettierten Beleg und verschiebt sich mit vertauschten Erfassungslabels, auch wenn tatsächliche Darbietungsfolge und Zeitabstände unabhängig kontrolliert werden.

Begründung. Die Darstellung könnte aus einer bloßen Ablagefolge einen sachlichen Geltungswechsel machen. Der eigenständige Gegenstand ist eine unbegründete Ersetzungsrelation, nicht allgemeine Bevorzugung neuer Informationen. Die vorliegenden Literaturberichte belegen verwandte Reihenfolge- und Aktualitätseffekte, aber keinen solchen Effekt irrelevanter Erfassungszeiten. Deshalb müssen Zeitetikett, tatsächliche Darbietung und allgemeine Verknüpfung getrennt werden. Die Hypothese beansprucht einen Darstellungsbefund; Aufmerksamkeits- und Lesemechanismen bleiben offen.

Falsifikationsbedingung. Für den vorab bestimmten Aufgabenbereich wird die Konjunktion entkräftet, wenn bei gültiger Durchführung und hinreichend engen Intervallen entweder ein praktisch relevanter zusätzlicher Anteil falscher Ersetzungsurteile oder deren vorhergesagte Bindung an das frühere Erfassungslabel ausgeschlossen wird. Mindestwirkungen werden vor der Hauptuntersuchung begründet. Ein breites Intervall, eine fehlende statistische Signifikanz oder eine bloß andere Endantwort entscheidet den Test nicht. Erklärt eine zeitlich bedeutungslose Verbindungskontrolle den gesamten Effekt ebenso gut, ist die spezifische Chronologiedeutung nicht gestützt.

Herkunft. warm_pick: empirie-chronologie-material — 06 Hypothesentag/2026-10-19. Literaturbefunde ausschließlich aus den bereitgestellten Dateien literatur-vorrecherche-2026-11-02.md und literatur-ergaenzung-2026-11-02.md; insbesondere Ecker et al. (2015), Autorenmanuskript https://www.emc-lab.org/uploads/1/1/3/6/113627673/ecker.2015.jml.pdf. Kein zusätzlicher Quellenzugriff durch diesen Prüfer.

Behobene Einwände. Erfassungslabels und wirkliche Darbietungszeiten werden im Kernsatz und im Entwurf getrennt. Der entscheidende Etikettenwechsel prüft die Richtung des zusätzlichen Ersetzungsfehlers. Allgemeine Verknüpfungssemantik und induzierte Aufmerksamkeit werden als konkrete Alternativen behandelt. Praktische Mindestwirkung, Präzision und abhängige Messungen bleiben explizite Planungsaufgaben statt erfundener Zahlen.

Prüfentwurf.

- **Material.** Vorab unabhängig geprüfte Aufgabenregeln machen eindeutig, dass alle Belege zum selben unveränderten Sachverhalt gelten und keiner einen anderen widerruft. Erfassungszeit bezeichnet nur Ablage, keine Messung, Revision oder Qualitätsstufe. Unterschiedlicher sachlicher Informationswert wird kontrolliert; Quellenunabhängigkeit allein genügt nicht.

- **Zuweisung.** Grafische zeitliche Verknüpfung versus unverknüpfte Übersicht und die Zuordnung früher/später Erfassungslabels werden unabhängig randomisiert. Eine kontrollierte Erstexposition mit ebenfalls unabhängig variierter Darbietungsreihenfolge prüft tatsächliche Aktualität. Danach stehen alle Belege in allen Bedingungen gleich lange gleichzeitig bereit. Die Präzisionsplanung muss diese Faktoren berücksichtigen.

- **Darstellungskontrolle.** Eine vorab entwickelte, möglichst geometrisch gleiche Verbindung ohne zeitliche Bedeutung begrenzt die Alternative allgemeiner Pfeil- oder Abhängigkeitssemantik. Position, Hervorhebung, Lesbarkeit und Informationsmenge werden abgeglichen. Eine vollkommen reine semantische Manipulation wird nicht vorausgesetzt; verbleibende Unterschiede werden berichtet.

- **Endpunkte.** Primär werden vorab regelbasiert kodierte falsche Ersetzungsurteile und die gerichtete Fehlerzuordnung zum jeweils früher etikettierten Beleg ausgewertet. Der verknüpfungsbedingte Zuwachs muss dem randomisierten Labelwechsel folgen; ein gleicher allgemeiner Labeleffekt in beiden Darstellungen erklärt diesen Zuwachs nicht.

- **Auswertung.** Alle zugewiesenen Personen bleiben in der primären Analyse. Personen- und Aufgabenabhängigkeit wiederholter Urteile wird berücksichtigt. Absolute Unterschiede samt Unsicherheitsintervallen, vorab begründete Mindestwirkungen und Sensitivität gegenüber tatsächlicher Expositionsreihenfolge werden berichtet. Verständnisprüfungen werden separat vorvalidiert; keine Auswahl nach späterem Missverständnis.

- **Grenze.** Ein durch die Grafik veränderter Blick- oder Leseverlauf kann Teil des Verfahrenseffekts sein. Ohne weitere Manipulation wird kein davon unabhängiger innerer Chronologiemechanismus behauptet. Die These gilt nicht für echte Rücknahmen oder veränderliche Sachverhalte.

H2 — Wer die Form zerbrechen darf

Kernsatz. In Schillers Lied von der Glocke lässt die Verflechtung von Meisterurteil, beauftragtem Formbruch und politischem Aufruhr gesellschaftliche Veränderung als Frage urteilsfähiger Leitung erscheinen: Nach dieser Lesart erhält angeleitetes Handeln ein anderes Gewicht als eigenmächtiger Ausbruch. Diese textinterne Autorisierung wird über die Verbindung von Können, Arbeitsauftrag und gesellschaftlichem Urteil hergestellt; ob sie über die sachgerechte Unterscheidung von Werkzuständen und Zeitpunkten hinausgeht, ist der entscheidende Prüfpunkt.

Begründung. Die notwendige Zerstörung der Gussform und der verurteilte Ausbruch eröffnen einen konkreten Textvergleich. Dabei sagt der Text kann, nicht darf; die Gesellen zerschlagen die Form auf Auftrag, und fertiges erkaltetes Werk sowie glühendes Erz sind verschiedene Zustände. Eine Zuständigkeitslesart müsste deshalb die Anweisungsautorität und ihre poetische Übertragung erklären, statt dem Meister exklusiv die ausführende Handlung zuzusprechen. Luserke-Jaqui hat die Verbindung von Meisterwarten und gesellschaftlicher Selbstbesinnung bereits beschrieben. Der engere Beitrag wäre die genaue Rekonstruktion des Übergangs von Können zu Zuständigkeit, nicht die Entdeckung der handwerklich-politischen Analogie. Eine Forschungslücke ist damit nicht belegt.

Falsifikationsbedingung. Die Lesart wird entkräftet, wenn eine Rekonstruktion der einschlägigen Übergänge und Gegenstellen die Verbindung von technischem Können, Anweisung und gesellschaftlicher Zuständigkeit nicht trägt oder wenn sachgerechte Arbeitsfolge und Materialzustände den Gesamtbefund mit weniger unbelegten Brückenannahmen erklären. Auch eine reine Folgenbewertung oder Warnung vor Kontrollverlust kann die Deutung verdrängen, wenn sie die Akteurs- und Modalstruktur ebenso vollständig erklärt. Das Wort kann allein belegt kein darf; politische Wertung allein belegt keine Zuständigkeitsordnung. Diese hermeneutische Entkräftung ist kein statistischer Nullhypothesentest.

Herkunft. zieldomaene Kulturphilosophie; cold_pick Schiller - Die Glocke als Regelungsgedicht und hylemorphische Form. Nachgereichte Quellenbefunde aus literatur-schiller-2026-11-02.md/.json: Schiller, Das Lied von der Glocke, Musen-Almanach für 1800, S. 251–272, besonders S. 267–269, https://de.wikisource.org/wiki/Das_Lied_von_der_Glocke_(1800); Luserke-Jaqui (2018), Das Lied von der Glocke (1800), S. 141–160, insbesondere S. 142–143, https://elibrary.narr.digital/xibrary/10.24053-9783772056512. Ergänzende Befunde zu Breu (2017) sowie nur Metadaten/Abstract zu Hahn (2016). Keine neue eigene Primär- oder Sekundärquellenlektüre dieses Prüfers.

Behobene Einwände. Kann und darf werden getrennt; normative Zuständigkeit wird als zu belegende Lesart behandelt. Anweisungsautorität des Meisters und ausführendes Handeln der Gesellen werden unterschieden. Erkaltetes fertiges Werk und glühendes Erz werden als verschiedene Zustände berücksichtigt. Luserke-Jaquis einschlägige Vorarbeit begrenzt den Neuheitsanspruch. Drei konkrete Konkurrenzlesarten prüfen denselben Textbestand, ohne der Hermeneutik ein ungeeignetes Experiment aufzuzwingen.

Prüfentwurf.

- **Methodentyp.** Vergleichende hermeneutische Prüfung einer textinternen Autorisierung; keine empirische Wirkungsstudie.

- **Textkorpus.** Alle einschlägigen Arbeits-, Formbruch- und Aufruhrpassagen samt Übergängen und Gegenstellen in der angegebenen historischen Textfassung. Die nachgereichte Vorrecherche bietet bereits überprüfte Ansatzstellen, aber keine vollständige Rekonstruktion.

- **Vergleichsraster.** Modalverben, Imperative, anweisende und ausführende Akteure, Adressierung, Sprechhaltung, Werkzustand, Zeitpunkt und Begründung der Wertung werden getrennt festgehalten. Der Meister wird als anweisender Akteur von den ausführenden Gesellen unterschieden. Sprecherwechsel werden nur angenommen, wenn der Text sie trägt.

- **Entscheidender uebergang.** Kann wird zunächst als Können gelesen. Zu begründen wäre, weshalb seine Einbettung in Arbeitsauftrag und politische Passage darüber hinaus gesellschaftliche Zuständigkeit modelliert. Die Abkühlung beziehungsweise das noch glühende Erz bleiben sachliche Alternativen zum behaupteten Normunterschied; sie dürfen nicht als gleichartige Fälle behandelt werden.

• **Konkurrenzlesarten.** Arbeitsfolgenlesart: Vollendung, Abkühlung und sachgerechter Zeitpunkt erklären den Gegensatz ohne zusätzliche gesellschaftliche Berechtigungsordnung.; Folgenlesart: Produktiver und zerstörerischer Formbruch unterscheiden sich durch ihr Ergebnis; die Akteursbindung ist hierfür nicht entscheidend.; Rhetorische Warnung: Die technische Analogie thematisiert Kontrollverlust, ohne gesellschaftliche Veränderungszuständigkeiten festzulegen.

• **Literarische vorarbeiten.** Die Verbindung von Meisterwarten und gesellschaftlicher Selbstbesinnung bei Luserke-Jaqui ist als Vorarbeit zu berücksichtigen, ebenso dessen Unterscheidung von Beschreibung und Programm. Befunde aus Breu und dem Hahn-Abstract werden nur innerhalb der dokumentierten Zugriffstiefe verwendet. Der engere Übergang zur Zuständigkeit bleibt eine Prüffrage ohne belegten Neuheitsanspruch.

• **Entscheidungsgrundlage.** Vollständigkeit und Tragfähigkeit der Textbelege sowie Zahl und Gewicht unbelegter Zusatzannahmen in den konkurrierenden Lesarten. Keine erfundenen Effektgrößen; Wortfrequenzen allein entscheiden die Deutung nicht.

• **Grenze.** Historische Leserwirkung, Autorenabsicht und allgemeine Rechte gesellschaftlicher Gruppen werden nicht aus der Textstruktur abgeleitet. Die vollständige hermeneutische Prüfung und die historische Neuheitsprüfung bleiben offen.

H3 — Vorleistung eröffnet einen gemeinsamen Zweck

Kernsatz. In offenen Kooperationen erhöht die zufällige Zuweisung zu einem Verfahren, das vor der gemeinsamen Zweckbildung eine kleine, freiwillige und tatsächlich nutzbare Vorleistung ermöglicht, gegenüber einem Verfahren mit entsprechender Absichtserklärung die Wahrscheinlichkeit eines tatsächlich begonnenen neuen gemeinsamen Vorhabens. Dieser Verfahrensvorteil wird zusammen mit innerhalb vorab begründeter Grenzen äquivalenten mittleren Zuverlässigkeitserwartungen der Beteiligten vorhergesagt; daraus folgt kein Ausschluss einer Vermittlung über Erwartungen.

Begründung. Der originelle Prüfgegenstand bleibt die Entstehung und tatsächliche Aufnahme eines zuvor von keinem Beteiligten allein verfolgten gemeinsamen Vorhabens. Vertrauen als produktive Bedingung neuer Ziele ist nach der bereitgestellten Literatur bereits vorgezeichnet. Der engere Beitrag liegt im Vergleich zweier praktisch durchführbarer Kooperationsverfahren bei gleicher Ausstattung am gemeinsamen Arbeitsbeginn. Weil freiwilliges Handeln nicht erzwungen randomisiert werden kann, bezieht sich die kausale Aussage auf die zugewiesene Gelegenheit und ihr Verfahren. Ob wirklich geleistete Vorarbeit den Effekt verursacht und welcher Vertrauensmechanismus beteiligt ist, sind nachgelagerte Fragen.

Falsifikationsbedingung. Die Konjunktion ist im registrierten Kooperationsbereich entkräftet, wenn bei validierter Durchführung ein hinreichend präzises Intervall einen praktisch relevanten Vorteil der Verfahrenszuweisung ausschließt oder ein Erwartungsunterschied außerhalb der vorab begründeten Äquivalenzgrenzen hinreichend präzise belegt wird. Für eine Stützung müssen sowohl der Kooperationsvorteil als auch Äquivalenz der vorab festgelegten Erwartungsmaße nachgewiesen werden. Ein nicht signifikanter Unterschied, ein Intervall über der Äquivalenzgrenze ohne klare Entscheidung oder ein nachträglich ausgewählter Teil tatsächlich vorleistender Paare genügt weder als Bestätigung noch allein als Widerlegung. Diese Aussage betrifft das zugewiesene Verfahren, nicht den isolierten Effekt ausgeführter Vorleistung.

Herkunft. devil_advocate-Kontrastprüfung der vollständigen Wochenläufe 05., 12. und 19.10.2026; Vault-Anker Luhmann - Vertrauen als Komplexitätsreduktion. Pragmatischer Anschluss Follett/Dewey gemäß bereitgestelltem Bericht zu Heenen und Brown (2026), https://doi.org/10.1017/psa.2026.10220. Weitere Abgrenzung über die bereitgestellten Literaturberichte zu Edmondson (1999), Falk/Kosfeld (2006) und Goeschl/Jarke (2017). Keine zusätzlichen Quellen- oder Vaultzugriffe durch diesen Prüfer.

Behobene Einwände. Der Kernsatz benennt jetzt denselben kausalen Verfahrenseffekt wie Randomisierung und primäre Analyse. Tatsächliche Freiwilligkeit wird als Durchführungseigenschaft geprüft, nicht mit Zufallszuweisung verwechselt. Erwartungsäquivalenz ist eine zusätzliche Ergebniskonstellation und kein Ausschluss eines Erwartungsmechanismus. Frühere Nutzungs- und Zeitvorteile werden trotz späterem Ressourcenausgleich ausdrücklich berücksichtigt. Entkräftung, fehlender Nachweis und unpräzise Ergebnisse werden getrennt.

Prüfentwurf.

- **Zielgroesse.** Primär der Intention-to-treat-Effekt der zufälligen Verfahrenszuweisung auf Paarebene, also unter Einschluss aller zugewiesenen Paare unabhängig davon, ob die Vorleistung tatsächlich erfolgt oder angenommen wird.

- **Aufgabe.** Eine offene Gestaltungsaufgabe ohne vorgegebenen gemeinsamen Zweck. Vor der Zuweisung werden individuelle Zweckvorstellungen erhoben. Ein neues gemeinsames Vorhaben muss nach vorab kodierbaren Kriterien von beiden individuellen Vorschlägen substanziell abweichen, von beiden angenommen und durch eine reale Folgehandlung begonnen werden. Kriterien und Kodierung werden unabhängig vorgeprüft; bloße Umbenennung oder Zusammenfügung bereits feststehender Einzelziele genügt nicht.

- **Vergleich.** Zufällige Zuweisung der Paare zu einem Verfahren mit freiwilliger realer Überlassung eines für das Gegenüber nutzbaren Beitrags oder zu einem Verfahren mit inhaltlich entsprechender Absichtserklärung ohne diese frühe Überlassung. Die konkrete Handlung wird nicht als freiwillig bezeichnet, wenn sie vorgeschrieben oder durch Zwang herbeigeführt wird. Eine unabhängige Pilotierung muss Umsetzbarkeit, Glaubwürdigkeit und Unterschiede in der tatsächlichen Nutzung belegen.

- **Zeitfolge.** Individuelle Ziele und Ausgangserwartungen; Zuweisung; freiwillige Vorleistung beziehungsweise Erklärung; erneute Erwartungserhebung; gemeinsame Zweckbildung; reale Folgehandlung. Die Erwartungserhebung erfolgt in allen Bedingungen gleich und vor der gemeinsamen Zielverhandlung.

- **Vergleichbarkeit.** Zum gemeinsamen Arbeitsbeginn vergleichbare Ausstattung und Handlungsmöglichkeiten. Informationsinhalt, Kontaktzeit und vorangegangene Bearbeitungszeit werden ebenfalls abgeglichen oder als verbleibende Unterschiede des Verfahrens offengelegt. Ressourcenausgleich allein beweist keine Gleichheit früherer Aufmerksamkeit oder Nutzungschancen.

- **Erwartungspruefung.** Vorab definierte Zuverlässigkeitsmaße für die Beteiligten werden nach Rollen getrennt zwischen den zugewiesenen Verfahren auf Äquivalenz geprüft; Ausgangswerte können präregistriert zur Präzisionssteigerung dienen. Grenzen werden aus Messgüte und sachlicher Relevanz begründet. Die gemeinsame Vorhersage umfasst alle vorab bezeichneten Erwartungsmaße. Auch nachgewiesene Äquivalenz mittlerer Werte ist keine Äquivalenz individueller Veränderungsverläufe und kein Nachweis fehlender Mediation.

- **Auswertung.** Absolute Unterschiede der begonnenen neuen Vorhaben mit Unsicherheitsintervallen; Präzisionsplanung auf Paarebene für Kooperationsvorteil und Erwartungsäquivalenz. Tatsächliche Vorleistungs- und Annahmequoten dokumentieren die Durchführung. Keine primäre Beschränkung auf tatsächlich Leistende oder Paare mit unveränderten Erwartungen. Ausfälle und fehlende Angaben werden nach einem vorab festgelegten Verfahren behandelt.

- **Alternative erklaerungen.** Gegenseitigkeitsverpflichtung nach empfangener Hilfe.; Selbstbindung der vorleistenden Person durch eigenen Aufwand.; Frühere Nutzbarkeit oder Bearbeitungserleichterung verändert die gemeinsame Ideengenerierung.

• **Grenze.** Kein isolierter kausaler Effekt tatsächlich freiwilliger Vorleistung, kein Beweis eines inneren Vertrauensmechanismus und keine Widerlegung systemtheoretischer Vertrauensbeschreibungen. Die endgültige Durchführbarkeit bleibt bis zur Pilotierung offen.

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	7	5	6
Falsifizierbarkeit	9	6	8
Begriffliche Klarheit	9	8	8
Tiefe	7	7	8
Forschungsrelevanz	8	7	8
Interdisziplinäre Anschlussfähigkeit	8	7	8
Vault-Anschluss	6	7	7
Antinomie-Test	7	7	7
Publikationsmöglichkeit	7	6	7
Summe /90	68	59	66

H1 — Die Chronologie erfindet einen Widerruf

Themenklasse: epistemologisch_systemtheoretisch_hypothese. **Methodentyp:** empirisch_pruefbar. **Reichweite:** these. Die Hypothese formuliert einen eigenständigen, begrenzten Zusammenhang zwischen zeitlicher Darstellung und falschem Geltungswechsel mit einem konkreten experimentellen Scheiterpunkt. Sie erweitert den dokumentierten Chronologie-Strang des Wochenlaufs vom 19. Oktober, ohne bereits drei dichte Vault-Knoten oder eine mehrgliedrige Forschungsarchitektur nachzuweisen.

Pflichttest: Grafische Verknüpfung, Zuordnung der Erfassungslabels und tatsächliche Darbietungsfolge sind unabhängig variierbar, falsche Ersetzungsurteile und deren Richtung regelbasiert messbar; damit ist der empirische Pflichtcheck auf Entwurfsebene bestanden, während Materialvalidierung, Mindestwirkungen und Präzisionsplanung noch ausstehen und keine zusätzlichen Höchstwertabsenkungen auslösen.

• **Originalität (7/10):** Die an irrelevante Erfassungslabels gebundene Zuschreibung eines nicht erfolgten Widerrufs grenzt einen eigenständigen Prüfgegenstand von bekannten Reihenfolge-, Aktualitäts- und echten Rücknahmeeffekten ab, dessen Neuheit die begrenzte Vorrecherche jedoch nicht sichert.

• **Falsifizierbarkeit (9/10):** Der unabhängig randomisierte Labelwechsel muss die Richtung des zusätzlichen Ersetzungsfehlers verschieben, während präzise ausgeschlossene Mindestwirkungen oder eine ausreichende zeitneutrale Verbindungserklärung die Vorhersage scheitern lassen, doch validierte

Aufgaben und begründete Testgrenzen fehlen noch.

• **Begriffliche Klarheit (9/10):** Der Entwurf trennt Erfassung, tatsächliche Exposition und sachliche Geltung ausdrücklich und definiert den behaupteten Fehler als zusätzliche, labelabhängige Ersetzung weiterhin gültiger Belege statt als bloß veränderte Endantwort.

• **Tiefe (7/10):** Der Übergang von einer Darstellungsordnung zu einem vermeintlichen Geltungswechsel berührt eine epistemische Grundunterscheidung, bleibt aber auf einen Darstellungseffekt ohne identifizierten inneren Mechanismus begrenzt.

• **Forschungsrelevanz (8/10):** Die Abgrenzung gegenüber den im Material dokumentierten Arbeiten von Ecker sowie Hey, Schaper und Bayen eröffnet einen konkreten Anschluss an die Forschung zu Erklärungsaktualisierung und fortwirkender Fehlinformation, ohne deren Befunde auf irrelevante Zeitlabels zu übertragen.

• **Interdisziplinäre Anschlussfähigkeit (8/10):** Der kontrollierte Übergang von visueller Ordnung zu Beleggeltung verbindet unmittelbar kognitive Psychologie, Informationsvisualisierung, Mensch-Computer-Interaktion und Erkenntnistheorie, während weitergehende Anwendungsfelder noch nicht ausgearbeitet sind.

• **Vault-Anschluss (6/10):** Der explizite Anschluss an 06 Hypothesentag/2026-10-19 trägt eine nachvollziehbare Fortsetzung des Chronologie-Strangs, doch der bereitgestellte Befund weist keine vertiefte Verbindung mehrerer dichter Sachknoten nach.

• **Antinomie-Test (7/10):** Die plausible Gegenthese einer additiven Belegverarbeitung ohne wirklichen Widerruf erhält durch Eckers nur schwache Evidenz gegen generelles automatisches Ersetzen eine begrenzte Gegenrichtung und wird durch den grafikspezifischen Labelwechsel gezielt prüfbar.

• **Publikationsmöglichkeit (7/10):** Der abgegrenzte Darstellungseffekt eignet sich für einen experimentellen Beitrag im Umfeld von Memory & Cognition, benötigt dafür jedoch validierte Materialien, ausreichende Präzision, Daten und eine vertiefte Neuheitsprüfung.

Stärkste Gegenthese: Erfassungszeiten ordnen weiterhin gültige unabhängige Belege, ohne deren additive Verarbeitung in eine Ersetzungsrelation zu verwandeln; zusätzliche Fehlurteile entstehen durch allgemeine Abhängigkeitssignale oder tatsächliche Aktualität und folgen deshalb keinem eigenständigen grafikspezifischen Wechsel der Erfassungslabels.

H2 — Wer die Form zerbrechen darf

Themenklasse: asthetisch_anschauliche_hypothese. **Methodentyp:** hermeneutisch. **Reichweite:** these. Die Einzelaussage betrifft den begrenzten poetischen Übergang von technischem Können und Arbeitsauftrag zu gesellschaftlicher Zuständigkeit und erweitert den benannten Schiller-Knoten. Ihr eigener Beitragswert ist durch konkurrierende Textrekonstruktionen prüfbar, während ein Forschungsprogramm oder eine historische Neuheitsbehauptung aus dem Material nicht folgt.

Pflichttest: Im bereitgestellten Quellenbefund sind Meisterurteil, beauftragter Formbruch und der Übergang von selbstbefreitem Erz zu politischem Aufruhr als drei konkrete Belegkomplexe benennbar, ergänzt durch Luserke-Jaqui und Breu; dies erfüllt die hermeneutische Mindestbasis, bestätigt aber weder den Übergang von kann zu darf noch die Vollständigkeit der Auslegung und führt zu keiner zusätzlichen Höchstwertabsenkung.

• **Originalität (5/10):** Meistervorbild, handwerklich-gesellschaftliche Analogie und sachgerechter Zeitpunkt sind nach Luserke-Jaqui und Breu bereits ausgelegt, sodass nur die noch unbewiesene Rekonstruktion einer besonderen Zuständigkeitsordnung zusätzlichen Beitragswert verspricht.

• **Falsifizierbarkeit (6/10):** Benannte Textübergänge und drei konkrete Konkurrenzlesarten erlauben eine quellengestützte Entkräftung durch geringeren Bedarf an unbelegten Brückenannahmen und erfüllen damit den hermeneutischen Höchstwert, ohne eine experimentelle Falsifikation vorzutäuschen.

• **Begriffliche Klarheit (8/10):** Können und Dürfen, Anweisung und Ausführung sowie erkaltetes Werk und glühendes Erz sind sauber getrennt, während textinterne Autorisierung erst durch die noch ausstehende Gesamtlektüre präzise ausgewiesen werden muss.

• **Tiefe (7/10):** Die Frage, wie poetisch dargestellte Sachkompetenz gesellschaftliche Veränderungszuständigkeit begründen kann, verbindet ästhetische Form und Autorität, bleibt aber von einem noch nicht belegten normativen Übergang abhängig.

• **Forschungsrelevanz (7/10):** Die Zuständigkeitslesart schließt konkret an die belegten Diskussionen über Meistervorbild, Ordnung und Revolutionskritik an, ohne dass bereits eine aktuell offene Forschungslücke nachgewiesen wäre.

• **Interdisziplinäre Anschlussfähigkeit (7/10):** Die Vermittlung zwischen Gussform, poetischer Analogie und gesellschaftlicher Autorität verbindet Literaturwissenschaft, Ästhetik und politische Philosophie, trägt aber noch keine allgemeine sozialwissenschaftliche Wirkungsbehauptung.

• **Vault-Anschluss (7/10):** Die Hypothese erweitert Schiller - Die Glocke als Regelungsgedicht und hylemorphische Form unmittelbar um die Frage der anweisenden Zuständigkeit, während die Dichte und der vorhandene Ausarbeitungsstand dieses Knotens im Bewertungsmaterial nicht dokumentiert sind.

• **Antinomie-Test (7/10):** Die Gegenthese, dass Werkreife, Arbeitsfolge und Folgen den Gegensatz vollständig erklären, ist wegen der unterschiedlichen Materialzustände stark und zwingt die Autorisierungslesart zu einem eigenständigen textlichen Nachweis.

• **Publikationsmöglichkeit (6/10):** Eine genaue Rekonstruktion des Übergangs von Können zu Zuständigkeit könnte einen literaturwissenschaftlichen Beitrag tragen, doch ohne vollständigen Textvergleich und Abgrenzung zur einschlägigen Schiller-Forschung bleibt der zusätzliche Erkenntnisgewinn offen.

Stärkste Gegenthese: Das Gedicht bewertet sachgerechte und verfrühte Formbrüche nach Werkzustand, Zeitpunkt und Folgen; Meisterurteil und Arbeitsauftrag sichern die technische Arbeitsfolge, ohne darüber hinaus eine besondere gesellschaftliche Berechtigung zur Veränderung zu begründen.

H3 — Vorleistung eröffnet einen gemeinsamen Zweck

Themenklasse: epistemologisch_systemtheoretisch_hypothese. **Methodentyp:** empirisch_pruefbar. **Reichweite:** these. Die Hypothese verbindet einen einzelnen kausalen Verfahrensvergleich mit einer ausdrücklich zusätzlichen Erwartungsäquivalenz und erweitert den benannten Vertrauensknoten durch einen operationalisierbaren Zielbildungsbefund. Die weiterreichenden Fragen nach Vertrauen, Reziprozität und gemeinsamer Zweckentstehung sind offene Anschlussfragen und bilden noch kein ausgearbeitetes Forschungsprogramm.

Pflichttest: Zufällige Verfahrenszuweisung, tatsächlich begonnene neue gemeinsame Vorhaben und rollenbezogene Zuverlässigkeitserwartungen lassen sich auf Paarebene erfassen; damit ist die empirische Operationalisierbarkeit im Entwurf gegeben, während Neuheitskodierung, Freiwilligkeit, Durchführbarkeit und Äquivalenzgrenzen erst unabhängig validiert werden müssen und keine zusätzliche Höchstwertabsenkung auslösen.

• **Originalität (6/10):** Gemeinsame Zielerzeugung durch vertrauensvolle Kooperation ist nach Heenen und Brown bereits vorgezeichnet, während der engere Verfahrensvergleich mit tatsächlichem Vorhabenbeginn und zusätzlicher Erwartungsäquivalenz einen möglichen, noch nicht als neu gesicherten

Beitrag bietet.

• **Falsifizierbarkeit (8/10):** Der Effekt der Verfahrenszuweisung und die Erwartungsäquivalenz bilden eine ausdrücklich scheiterbare Konjunktion ohne Auswahl tatsächlich Leistender, doch ihre Prüfschärfe hängt noch von belastbarer Neuheitskodierung, Durchführung und sachlich begründeten Äquivalenzgrenzen ab.

• **Begriffliche Klarheit (8/10):** Verfahrenszuweisung, freiwillige Handlung und Erwartungsmediation werden klar unterschieden, während die substantielle Neuheit eines gemeinsamen Zwecks und die Vergleichbarkeit früher Nutzungschancen noch konkrete validierte Kriterien benötigen.

• **Tiefe (8/10):** Die Entstehung zuvor nicht einzeln verfolgter gemeinsamer Zwecke berührt die Konstitution kollektiven Handelns, während der begrenzte Verfahrenseffekt weder einen inneren Vertrauensmechanismus noch eine allgemeine Theorie der Zweckbildung beweist.

• **Forschungsrelevanz (8/10):** Der Vorschlag verbindet den aktuellen Follett-Dewey-Anschluss bei Heenen und Brown mit den bereitgestellten Befunden zu psychologischer Sicherheit und Kontrolle, ohne diese verschiedenen Konstrukte gleichzusetzen.

• **Interdisziplinäre Anschlussfähigkeit (8/10):** Gemeinsame Zweckbildung unter einer randomisierten Kooperationsordnung ist unmittelbar für Sozialpsychologie, experimentelle Ökonomik, Organisationsforschung und pragmatistische Sozialphilosophie anschlussfähig.

• **Vault-Anschluss (7/10):** Der direkte Bezug auf Luhmann - Vertrauen als Komplexitätsreduktion und die dokumentierte Kontrastprüfung früherer Wochenläufe tragen eine produktive Ergänzung um neu entstehende Zwecke, belegen aber weder einen Widerspruch zu Luhmann noch mehrere ausgearbeitete dichte Knoten.

• **Antinomie-Test (7/10):** Die Gegenthese, dass frühe Vorleistung durch Reziprozitätsdruck und Festlegung den Raum für neue gemeinsame Zwecke verengt, ist ebenso plausibel wie die Öffnungsthese und lässt sich am selben Verfahrensvergleich entscheiden.

• **Publikationsmöglichkeit (7/10):** Der kausale Vergleich mit einem beobachtbaren Vorhabenbeginn ist als sozialwissenschaftlicher Experimentalbeitrag tragfähig, benötigt jedoch eine erfolgreiche Pilotierung, validierte Messungen und einen klareren eigenständigen Beitrag gegenüber der vorhandenen Vertrauensliteratur.

Stärkste Gegenthese: Ein Verfahren mit der Gelegenheit zu früher realer Vorleistung verstärkt Festlegung und Gegenseitigkeitsverpflichtung stärker als die offene Suche nach neuen Zwecken; ein Verfahren mit entsprechender Absichtserklärung lässt deshalb mehr Raum für eigenständig neue gemeinsame Vorhaben und begünstigt deren tatsächlichen Beginn mindestens ebenso stark.

Auswahl und Qualitätsentscheid

H1: Die Chronologie erfindet einen Widerruf (68/90) → H3: Vorleistung eröffnet einen gemeinsamen Zweck (66/90) → H2: Wer die Form zerbrechen darf (59/90).

H1 erreicht 68 Punkte vor H3 mit 66 und H2 mit 59 Punkten und geht aufgrund der präzisen Trennung von Erfassungslabel, Darbietungsfolge und zusätzlichem Ersetzungsurteil in die Expertenrunde. Die beiden 9er-Werte für Falsifizierbarkeit und begriffliche Klarheit beruhen auf der konkreten Logik des randomisierten Labelwechsels und der ausdrücklich getrennten Geltungs- und Zeitbegriffe, nicht auf einer empirischen Bestätigung. Eckers begrenzter Befund gegen ein allgemeines automatisches Ersetzen ohne Rücknahme verschärft die plausible Gegenthese, prüft aber H1s grafische Verknüpfung und irrelevante Erfassungslabels nicht. Für Originalität, Tiefe und Publikationsmöglichkeit ist bei keinem Kandidaten ein 9er- oder 10er-Score hinreichend belegt, weil die Neuheitsprüfung begrenzt bleibt,

empirische Pilotierung beziehungsweise vollständige Textrekonstruktion fehlen und keine hohe Annahmewahrscheinlichkeit begründet werden kann. Die gewichteten Summen sind relative Ordnungsurteile, sodass der geringe Abstand zwischen H1 und H3 keine statistisch gesicherte Qualitätsdifferenz bedeutet.

Hauptlinie: Die Chronologie erfindet einen Widerruf. **Nebenlinie:** Vorleistung eröffnet einen gemeinsamen Zweck. 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 19.10. eingehalten.

Expertenrunde 1 — Hauptlinie

Hypothese. Die Chronologie erfindet einen Widerruf (68/90).

Kernsatz. Bei weiterhin gültigen, sachlich unabhängigen Belegen zu einem unveränderten Sachverhalt erhöht eine grafische Verknüpfung nach sachlich irrelevanten Erfassungszeiten gegenüber einer informationsgleichen unverknüpften Übersicht falsche Ersetzungsurteile. Der zusätzliche Fehler richtet sich gegen den jeweils früher etikettierten Beleg und verschiebt sich mit vertauschten Erfassungslabels, auch wenn tatsächliche Darbietungsfolge und Zeitabstände unabhängig kontrolliert werden.

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

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

Gutachten 1 — Kant

Die Bedingungen der Möglichkeit eines Ersetzungsurteils liegen nicht schon in der Ordnung des Nacheinanders: Eine zeitliche Relation verbindet Vorstellungen, während ein Widerruf eine Regel verlangt, nach welcher einem Beleg seine bisherige Geltung entzogen wird. Die Hypothese gewinnt ihre Bestimmtheit dort, wo sie diese zweite Relation als unbegründete Zutat der Urteilskraft untersucht. Dabei ist die im Versuch gesetzte Regel der fortbestehenden Geltung der Maßstab des Fehlers; sie darf nicht ihrerseits aus dem beobachteten Antwortverhalten gewonnen werden. Mein Einwand betrifft jedoch die noch unvollständige Bestimmung des zu messenden Urteilsakts: Ein weiterhin gültiger Beleg kann für eine Entscheidung entbehrlich werden, ohne dadurch widerrufen zu sein. Wer den früheren Beleg nicht mehr anführt oder den späteren als ausreichend bezeichnet, hat deshalb noch keinen Geltungswechsel behauptet. Auch eine vorab festgelegte Kodierung beseitigt diese Schwierigkeit nur, wenn sie die Unterscheidung sachlich abbildet. Die Schärfung sollte daher im Antwortformat erfolgen: Fortbestehende Geltung bei fehlender zusätzlicher Entscheidungsrelevanz muss von entzogener Geltung ausdrücklich unterscheidbar sein; eine kurze Begründung soll erkennen lassen, welche Regel den behaupteten Entzug tragen soll. Diese Unterscheidung ist unabhängig vorzuprüfen, ohne später Personen nach ihrer Antwort auszuschließen. Erst wenn der grafikspezifische Labelwechsel gerade die unbegründete Entziehung der Geltung verschiebt, trifft der Befund den angekündigten Gegenstand. Schließlich wäre damit eine empirische Verwechslung von Darstellungsordnung und Begründungsordnung gezeigt; weder die zeitliche Form überhaupt noch das Vermögen vernünftiger Beurteilung wäre dadurch als fehlerhaft erwiesen.

Gutachten 2 — Popper

Die riskante Vorhersage liegt nicht im bloßen Anstieg falscher Antworten, sondern darin, dass der zusätzliche Fehler mit einer sachlich bedeutungslosen Etikettenvertauschung seine Richtung wechselt. Das macht aus einer allgemeinen Warnung vor Zeitachsen eine prüfbare Vermutung. Mein Einwand betrifft jedoch den logischen Status der Verbindungskontrolle: Dass eine zeitneutrale Verbindung den Effekt ebenso gut erklärt, ist ohne festgelegtes Vergleichskriterium noch keine entscheidbare Bedingung. Hier könnte ein unerwünschtes Ergebnis nachträglich zur bloßen Unklarheit erklärt werden. Mein Rat: Bestimmen Sie vorab auch den direkten Kontrast zwischen zeitlicher und zeitneutraler Verknüpfung einschließlich seiner gerichteten Bindung an die Erfassungslabels. Ein präzise ausgeschlossener Mindestunterschied muss die spezifische Chronologiebehauptung zu Fall bringen dürfen, selbst wenn beide Grafiken gegenüber der unverknüpften Übersicht Fehler erzeugen. Die Mindestwirkung dieses spezifischen Kontrasts braucht eine eigene sachliche Begründung; sie darf nicht erst aus dem beobachteten Gesamteffekt abgeleitet werden. Außerdem muss die Kodierung eine logische Unterscheidung sichern: Einen älteren Beleg nicht zu verwenden, ihn geringer zu gewichten oder ihn für redundant zu halten, behauptet noch keinen Widerruf. Erfasst werden sollte deshalb eine ausdrückliche Aberkennung seiner weiterhin bestehenden Geltung, mit vorvalidierten Fragen, die diese Antwort nicht bereits nahelegen. Für ungültige Durchführung sind Ausschlussbedingungen auf Studienebene vorab festzulegen; nachträgliches Unverständnis einzelner Personen darf die These nicht retten. Diese Schärfungen verlangen keinen Nachweis eines inneren Mechanismus. Sie verhindern vielmehr, dass eine bescheidene, falsifizierbare Darstellungsbehauptung durch wechselnde Erfolgskriterien gegen Widerlegung geschützt wird.

Gutachten 3 — Wittgenstein

Wer sagt, ein Beleg sei ersetzt, kann Verschiedenes tun. Er kann ihm die Geltung absprechen, ihn bei einer Antwort weglassen oder die Bearbeitung seiner Akte beenden. Diese Verwendungen sind nicht austauschbar. Der Entwurf unterscheidet Erfassungszeit und sachliche Geltung sorgfältig. Noch entscheidet diese Unterscheidung aber nicht, welche Sprachhandlung eine Versuchsperson mit ihrer Antwort vollzieht. Stellen Sie sich vor, jemand sagt angesichts zweier gültiger Nachweise: Für meinen Bericht nehme ich den später abgelegten. Er könnte den früheren weiterhin für gültig halten und lediglich eine Auswahl treffen. Selbst die Wendung „der neue ersetzt den alten“ kann eine solche Arbeitsregel ausdrücken. Mein Einwand betrifft deshalb die Übersetzung von Antworten in Ersetzungsurteile: Eine vorab festgelegte Kodierung schützt nicht davor, einen mehrdeutigen Gebrauch einheitlich falsch zu lesen. Zeigen Sie anhand konkreter Antwortfälle, woran Geltungsentzug erkennbar wird. Fragen Sie getrennt, ob der frühere Beleg weiterhin zur Begründung derselben Aussage verwendet werden darf und ob jemand seine Aussage zurückgenommen hat. Halten Sie daneben fest, welchen Beleg die Person tatsächlich auswählt. Diese Unterscheidungen müssen in den Aufgabenbeispielen verständlich werden, ohne den Teilnehmenden die vermutete Fehlerquelle zu verraten. Dabei ist ein behaupteter Widerruf enger als jeder Verlust zugeschriebener Geltung. Die Grafik könnte etwa eine administrative Auswahl nahelegen. Erst wenn sich die Antworten gegen die fortbestehende Verwendbarkeit richten, trifft die Diagnose der unbegründeten Ersetzung. Die Chronologie widerruft nichts. Untersucht wird, unter welchen Darstellungsbedingungen Menschen von einem weiterhin gültigen Beleg anders Gebrauch machen.

Gutachten 4 — Plessner

Die grafische Ordnung begegnet dem Urteilenden als bereits gestaltete Weise, in der ein anderer ihm einen Sachverhalt erschließt. Gerade darin liegt die anthropologische Pointe: Der Leser kann die Darstellungsregel ausdrücklich kennen und sich dennoch an einer vermuteten Mitteilungsabsicht orientieren. Eine Verbindung zwischen zwei Belegen kann heißen: So sollst du ihren Zusammenhang verstehen. Der behauptete Fehler wäre dann eine Verschiebung zwischen der Stellung zum dargestellten Sachverhalt und der Stellung gegenüber einer als sinnvoll vorausgesetzten Darstellung. Der Entwurf schützt diese Unterscheidung durch seine Beschränkung auf einen Verfahrenseffekt; er sollte sie auch in der Deutung bewahren. Mein Einwand lautet, dass die Grafik zugleich eine soziale Erwartung an den Antwortenden vermitteln kann: Wer den früheren Beleg ausschließt, könnte eine vermutete Prüfungsanforderung erfüllen, ohne dessen Geltung selbst zu bestreiten. Der kontrollierte Labelwechsel wäre eine notwendige Bedingung für die beanspruchte Fehlerzuordnung, keine hinreichende Beschreibung dieses Erfahrungsunterschieds. Zur Schärfung empfehle ich deshalb eine nachgelagerte, getrennt ausgewertete Erhebung: Teilnehmende sollen knapp erläutern, ob sie einen sachlichen Widerruf, eine bloße Empfehlung der Darstellung oder eine erwartete Antwort annahmen. Dabei bleiben alle Personen in der primären Analyse; die Erläuterungen dürfen den experimentellen Befund weder nachträglich bereinigen noch seinen Mechanismus beweisen. Eine unabhängig pilotierte Variante könnte die Grafik ausdrücklich einer automatischen Ablagesortierung zuschreiben. So würde prüfbar, ob bereits die zeitliche Form trägt oder ob ihre Wirkung an die vermutete Urteilskompetenz eines Darstellenden gebunden ist.

Gutachten 5 — Cassirer

Die behauptete Verwandlung einer Ablagefolge in einen Geltungswechsel betrifft die Tätigkeit der Darstellung selbst: Eine grafische Verbindung stellt nicht lediglich bereits fertige Erkenntnisse nebeneinander, sondern bietet eine Regel an, unter der ihre Beziehungen verstanden werden. Aus dieser Perspektive ist der interessante Fehler ein Wechsel zwischen zwei Ordnungsfunktionen: Die Zeitangabe bestimmt zunächst den Platz eines Belegs im Archiv und übernimmt anschließend unberechtigt die Bestimmung seines sachlichen Ranges. Der Entwurf gewinnt dadurch, dass er diesen Übergang gegenüber tatsächlicher Aktualität und berechtigtem Widerruf abgrenzt. Mein Einwand richtet sich jedoch gegen eine mögliche Gleichsetzung von Informationsgleichheit und gleicher Bedeutung. Identische Sätze und Daten garantieren keine identische symbolische Funktion; gerade deren Veränderung soll das Experiment untersuchen. Auch die Irrelevanz einer Erfassungszeit gilt zunächst innerhalb der Aufgabenregel, während eine vertraute grafische Konvention weiterhin Revision, Fortschritt oder Ablösung nahelegen kann. Folgt der Fehler dem vertauschten Zeitlabel, wäre damit der Darstellungsbefund gestützt, aber noch nicht entschieden, welche dieser Bedeutungen wirksam wurde. Zur Schärfung sollte deshalb vorab an gesondertem Material geprüft werden, welche Beziehungen die verwendeten Zeichen für die Zielgruppe ausdrücken; nach dem primären Urteil könnte eine kurze, vorab kodierte Erläuterung zwischen Ablagefolge, sachlicher Abhängigkeit und Ersetzung unterscheiden. Diese Erläuterungen wären ergänzende Deutungsdaten, keine nachträglichen Ausschlussgründe und kein Beweis eines inneren Mechanismus. So würde die Chronologie nicht selbst zum handelnden Wesen erhoben: Untersucht würde die konkrete symbolische Leistung, durch die eine Ordnung des Auffindens als Ordnung der Geltung gelesen wird.

Gutachten 6 — Friston

Aus bayesianischer Perspektive kann der behauptete Fehler eine intern kohärente Inferenz unter einem unpassenden generative model sein: Die Versuchsperson behandelt eine grafisch geordnete Belegfolge als Abfolge von Zustandsrevisionen, obwohl die Aufgabenwelt ausschließlich zusätzliche Beobachtungen desselben Zustands enthält. Das frühere Etikett würde dann nicht unmittelbar einen Beleg entwerten; es bestimmt dessen Stellung in einem fälschlich angenommenen Übergangsmodell. Diese Unterscheidung ist mathematisch erheblich, denn das freie Energieprinzip allein sagt keinen Ersetzungsfehler voraus. Dafür müssten Hypothesenraum, Übergangswahrscheinlichkeiten und die zugeschriebene Verlässlichkeit der Grafik spezifiziert werden. Der vorliegende Entwurf identifiziert überzeugend einen möglichen Verfahrenseffekt, identifiziert jedoch kein solches Modell. Mein konkreter Einwand betrifft deshalb die Deutung des Ersetzungsurteils: Dieselbe Antwort könnte aus einer veränderten Geltungsannahme oder aus einer Antwortstrategie entstehen, die den vermeintlich erwarteten neuesten Beleg auswählt. Der Labelwechsel trennt diese Möglichkeiten noch nicht. Als begrenzte Schärfung empfehle ich eine getrennte Validierungsstudie, in der die fortbestehende Gültigkeit jedes Belegs unabhängig von dessen Auswahl für eine Abschlussantwort beurteilt wird; Antwortpositionen und Antwortformulierungen sollten dabei ausbalanciert sein. Vorab formulierte Modelle einer Geltungsrevision und einer bloßen Auswahlpräferenz könnten anschließend an ihren unterschiedlichen Vorhersagen verglichen werden. Auch solche Urteile wären keine unmittelbaren Fenster auf innere Zustände, würden aber zusätzliche Identifikationsbedingungen liefern. Für den Hauptversuch bleibt der vorsichtige Anspruch eines Darstellungseffekts angemessen. Erst eine erfolgreiche Modellunterscheidung rechtfertigte die stärkere Aussage, dass die Chronologie tatsächlich einen inneren Widerruf erzeugt, statt lediglich dessen sprachlichen Ausdruck zu begünstigen.

Gutachten 7 — Luhmann

Die Hypothese untersucht eine mögliche Umstellung der Unterscheidung: Aus früher/später wird gilt/gilt nicht mehr. Auffällig ist dabei weniger die Zeitordnung selbst als die kommunikative Zurechnung einer Berechtigung zur Revision. Ein früheres Datum könnte als Signal eines überholten Standes behandelt werden, wenn der Beobachter der Darstellung einen Aktualisierungsanspruch zuschreibt. Diese Bedingung ist theoretisch plausibel, durch den vorgesehenen Verfahrensvergleich aber noch nicht identifiziert. Der konkrete Einwand lautet daher: Das Experiment bietet zugleich die ausdrückliche Mitteilung fortdauernder Gültigkeit und eine möglicherweise als Statusmitteilung verstandene grafische Ordnung an; ein Ersetzungsurteil kann anzeigen, welcher Mitteilung die Versuchsperson Autorität zuschreibt. Der zusätzliche Fehler wäre damit eine beobachtbare Auswahl zwischen widersprüchlich gelesenen Kommunikationen, ohne dass bereits entschieden wäre, ob Zeitordnung als solche Geltungsordnung erzeugt. Die vorhandene zeitneutrale Verbindungskontrolle begrenzt diesen Einwand, beseitigt ihn jedoch nicht, weil auch der Aktualisierungsanspruch gerade an Datierungen anschließen könnte. Zur Schärfung sollte die Untersuchung nach den primären Urteilen getrennt erfassen, ob der frühere Beleg für sachlich falsch, weiterhin gültig, aber entbehrlich, oder durch eine zuständige Instanz zurückgenommen gehalten wird. Diese Antworten dienen der Beschreibung möglicher Zurechnungen und dürfen keine nachträgliche Auswahl der untersuchten Personen begründen. Eine spätere Anschlussstudie könnte bei identischen Darstellungen die zugeschriebene Funktion der Oberfläche variieren: bloßes Ablageverzeichnis oder redaktionell verantworteter Geltungsstand. Systemtheoretisch interessant wäre dann die Bedingung, unter der eine Ordnung der Speicherung als Kommunikation über Verbindlichkeit behandelt wird. Der eng begrenzte Darstellungseffekt bleibt unabhängig davon prüfbar; seine theoretische Erklärung gewinnt erst durch die Beobachtung der verwendeten Unterscheidungen Kontur.

Nebenlinie — sieben Kurzbegutachtungen

Hypothese. Vorleistung eröffnet einen gemeinsamen Zweck (66/90).

Kernsatz. In offenen Kooperationen erhöht die zufällige Zuweisung zu einem Verfahren, das vor der gemeinsamen Zweckbildung eine kleine, freiwillige und tatsächlich nutzbare Vorleistung ermöglicht, gegenüber einem Verfahren mit entsprechender Absichtserklärung die Wahrscheinlichkeit eines tatsächlich begonnenen neuen gemeinsamen Vorhabens. Dieser Verfahrensvorteil wird zusammen mit innerhalb vorab begründeter Grenzen äquivalenten mittleren Zuverlässigkeitserwartungen der Beteiligten vorhergesagt; daraus folgt kein Ausschluss einer Vermittlung über Erwartungen.

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

Die Stärke der Nebenlinie liegt in der Trennung zwischen gewährter Handlungsgelegenheit und tatsächlich freiwilliger Handlung; damit bleibt Freiheit wenigstens begrifflich von der Verfahrenszuweisung unterschieden. Ihre Schwäche betrifft die Bestimmung des gemeinsamen Zwecks: Inhaltliche Neuheit, beiderseitige Annahme und ein beobachtbarer Beginn sichern noch keine gemeinsame Urheberschaft, denn auch ein aus empfundener Schuld übernommener Vorschlag kann diese Kriterien erfüllen. Forschungspotenzial eröffnet eine ergänzende Prüfung, ob beide Beteiligten den Zweck mit eigenen Gründen vertreten und Änderungswünsche wirksam einbringen können, ohne die Zustimmung als geschuldete Gegenleistung zu behandeln. Solche Beobachtungen wären begrenzte Hinweise auf die Gestalt der Zweckbildung, kein empirischer Beweis autonomer Willensbestimmung. Der bisherige Verfahrensvergleich bliebe tragfähig; seine philosophische Deutung müsste zwischen erfolgreicher Koordination und gemeinsam begründeter Zwecksetzung unterscheiden.

Kurzbegutachtung 2 — Popper

Die Stärke liegt in der ehrlichen Beschränkung auf das zugewiesene Verfahren: Freiwillige Vorleistung wird nicht nachträglich wie eine randomisierte Eigenschaft behandelt. Die Schwäche sitzt im Neuheitsbegriff. Aus nicht erhobenen Einzelabsichten folgt nicht, dass ein Zweck zuvor von niemandem verfolgt wurde; selbst sorgfältige Erstbefragungen liefern keinen vollständigen Katalog möglicher Ziele. Begrenzen Sie die Behauptung daher auf substanzielle Abweichungen von den dokumentierten Einzelzielen und lassen Sie diese anhand festgelegter Kriterien ohne Kenntnis der Versuchsbedingung beurteilen. Forschungspotenzial bietet die getrennte Beobachtung von gemeinsamer Zielannahme und tatsächlichem Beginn: Vorleistung könnte Einigung erleichtern, zugleich aber die Umsetzung durch Verpflichtungsdruck hemmen. Die vorgeschlagene Konjunktion bleibt prüfbar; ähnliche mittlere Erwartungen erklären allerdings weder diesen Übergang noch schließen sie individuelle Erwartungsänderungen aus.

Kurzbegutachtung 3 — Wittgenstein

Die Stärke liegt im tatsächlichen Beginn eines gemeinsamen Vorhabens: Die Beteiligten müssen mit ihrem Vorschlag etwas anfangen. Ein freundliches Einverständnis genügt nicht. Die Schwäche steckt im Ausdruck „zuvor von keinem allein verfolgt“. Eine vorab geäußerte Zweckvorstellung kann ein Wunsch, ein Arbeitsauftrag oder ein ausgearbeiteter Plan sein. Derselbe spätere Vorschlag erscheint je nach Ausführlichkeit dieser ersten Äußerung neu oder bloß genauer. Die Kodierung könnte damit sprachliche Ausführlichkeit als Zweckentstehung behandeln. Forschungspotenzial bietet eine Sammlung konkreter Grenzfälle: Wann nennen Beteiligte eine Änderung bloß eine Präzisierung, wann geben sie dafür einen bisherigen Zweck auf? Solche Fälle könnten die Neuheitskriterien schärfen. Daraus folgt weder, dass gemeinsame Zwecke nur Benennungen wären, noch dass eine Vorleistung bereits Vertrauen ausdrückt.

Kurzbegutachtung 4 — Plessner

Die Stärke des Vorschlags liegt darin, dass eine Vorleistung den anderen als jemanden anspricht, dessen Antwort noch offensteht. Ein gemeinsamer Zweck kann in dieser vermittelten Begegnung entstehen, weil beide ihre zunächst eigene Stellung verändern. Die Schwäche ist die Gleichsetzung kodierter Zweckneuheit mit solcher Gegenseitigkeit: Auch ein unter Dankesschuld übernommener Vorschlag kann von beiden Ausgangszielen abweichen und praktisch begonnen werden. Die vorgesehenen Alternativerklärungen sollten deshalb um die erlebte Möglichkeit ergänzt werden, einen Beitrag abzulehnen oder umzudeuten. Forschungspotenzial bietet eine begleitende Rekonstruktion konkreter Aushandlungen: Wer verändert wessen Vorschlag, welche Einwände erhalten Raum, und wird ein Nein akzeptiert? Solche Beschreibungen ergänzen den kausalen Verfahrensvergleich; äquivalente Zuverlässigkeitserwartungen erfassen diese wechselseitige Stellung nicht vollständig.

Kurzbegutachtung 5 — Cassirer

Die Stärke der Nebenlinie liegt darin, den gemeinsamen Zweck als Ergebnis einer Tätigkeit zu behandeln: Eine Vorleistung kann einen Gegenstand gemeinsamer Bearbeitung eröffnen, dessen Bedeutung vor der Begegnung noch nicht feststand. Ihre Schwäche besteht darin, dass Nutzbarkeit und symbolische Bedeutung des Beitrags auseinanderfallen können. Dieselbe Gabe kann als Einladung, als Anspruch auf Gegengabe oder als Festlegung des zulässigen Vorhabens verstanden werden; vergleichbare Ausstattung und mittlere Zuverlässigkeitserwartungen entscheiden darüber nicht. Forschungspotenzial bietet deshalb eine ergänzende Untersuchung der gemeinsamen Umdeutung des Beitrags: Dürfen beide Beteiligten seinen Verwendungszweck verändern, und wird diese Veränderung tatsächlich aufgegriffen? Vorab bestimmte Beobachtungskriterien für Vorschläge, Widerspruch und angenommene Umwidmung könnten den experimentellen Verfahrensvergleich präzisieren, ohne daraus bereits einen nachgewiesenen Vertrauensmechanismus abzuleiten.

Kurzbegutachtung 6 — Friston

Die Stärke liegt darin, gemeinsame Zweckbildung als Veränderung des verfügbaren Handlungsraums zu untersuchen: Eine nutzbare Vorleistung könnte bisher unbekannte gemeinsame policies sichtbar machen, selbst wenn die mittlere Einschätzung persönlicher Zuverlässigkeit gleich bleibt. Die Schwäche ist die theoretisch noch unbestimmte Verbindung zwischen Erwartungsäquivalenz und diesem Vorgang. Gleiche Mittelwerte beschränken weder die Unsicherheit über Erwartungen noch die angenommene Umsetzbarkeit gemeinsamer Handlungen; daher entsteht aus ihrer Äquivalenz kein besonderer Widerspruch zu einem bayesianischen Kooperationsmodell. Forschungspotenzial bietet eine ergänzende Prozessstudie, die vor und nach der Vorleistung das wahrgenommene Repertoire gemeinsamer Möglichkeiten sowie die Sicherheit dieser Einschätzungen erfasst. Diese Größen sollten konkurrierende Mechanismen erschließen, ohne den randomisierten Verfahrenseffekt durch nachträgliche Auswahl oder Adjustierung tatsächlich vorleistender Paare umzudeuten.

Kurzbegutachtung 7 — Luhmann

Die Stärke der Nebenlinie liegt darin, den gemeinsamen Zweck als Ergebnis anschließender Kommunikation zu behandeln: Die Vorleistung schafft einen Bezugspunkt, an dem Zustimmung, Ablehnung und Gegenseitigkeit erst bestimmbar werden. Die Schwäche betrifft die Reichweite: Die verlangte inhaltliche Neuheit erfasst nur einen Sonderfall gemeinsamer Zweckbildung, denn auch ein vorab bekannter Zweck kann erst im Austausch kollektiv verbindlich werden. Forschungspotenzial bietet deshalb die getrennte Beobachtung von Zweckinnovation und kommunikativer Bindung: Wer nimmt den Beitrag als Angebot auf, wer formuliert eine Anschlusszusage, und welche Folgehandlung bestätigt diese? Äquivalente Zuverlässigkeitserwartungen wären damit vereinbar, weil psychische Einschätzungen und die Fortsetzung einer Kooperation unterschiedliche Beobachtungsgegenstände bleiben.

Nebenlinie — getrennte Quellenhinweise und offene Fragen

Vorleistung eröffnet einen gemeinsamen Zweck, 66/90. Sieben Kurzbegutachtungen abgeschlossen; keine zweite Runde. Die folgenden Quellenhinweise betreffen ausschließlich diese Nebenlinie.

- Armin Falk; Michael Kosfeld (2006): The Hidden Costs of Control. Im Prinzipal-Agenten-Experiment kann eine kontrollierende Mindestanforderung die Leistung mindern; Befragte interpretieren sie unter anderem als Misstrauen und Autonomiebeschränkung. **Grenze:** Kontrollverzicht als solcher ist keine neue Entdeckung; konkrete Übertragung auf zeitlich begrenzte Prüfaussetzung bleibt offen. **Lektüre:** Originalabstract im AEA-Websuchindex heute; Direktaufruf403. IZA-Fassung als weiterer Quellenhinweis, nicht zusätzlich ausgewertet.

- Timo Goeschl; Johannes Jarke (2017): Trust, but verify? Monitoring, inspection costs, and opportunism under limited observability. Ein wiederholtes Vertrauensspiel variiert die Beobachtbarkeit und Kosten aktiver Kontrolle. Vertrauen und Effizienz unterscheiden sich laut Abstract nicht signifikant über Bedingungen; teure Kontrolle geht mit vielen unbeobachteten Vertrauenshandlungen einher. **Grenze:** Kein Nachweis höherer Kooperation durch Nichtprüfung, kein Test eines unmittelbar nachfolgenden Prüfanspruchs. Nullbefund ist weder Äquivalenzbeweis noch allgemeine Wirkungsfreiheit. **Lektüre:** Original-Verlagsauszug im Websuchindex; Direktzugriff403. Autorschaft und Veröffentlichung am Lehrstuhl Heidelberg bestätigt. Vollständiges Abstract zusätzlich im bibliographischen RePEc-Datensatz; kein Volltext.

- Amy Edmondson (1999): Psychological Safety and Learning Behavior in Work Teams. Die Feldstudie mit 51 Arbeitsteams verbindet psychologische Sicherheit mit Lernverhalten. **Grenze:** Beobachteter Zusammenhang statt randomisiertem Vorleistungsvergleich; kein Test neu gemeinsam gebildeter Zwecke. Psychologische Sicherheit ist keine einzelne Vorleistung. **Lektüre:** Originalabstract und Metadaten im Harvard-Repositorium geöffnet; kein Volltext gelesen.

- Sunny Heenen; Matthew J. Brown (2026): Public Trust of Experts in Democratic Societies: Lessons from Follett and Dewey. Nicht hierarchische Zusammenarbeit und Vertrauen werden ausdrücklich mit gemeinsamer Entwicklung von Wissen, Zielen und Lösungen verbunden. **Grenze:** Der allgemeine Gedanke kooperativer Zweckbildung ist dadurch keine neue Entdeckung. H3s enges randomisiertes Verfahren wird nicht untersucht; Neuheit dieses Designs ist ebenfalls nicht bewiesen. **Lektüre:** Verlagsvolltext geöffnet, Abstract, Einleitung und Abschnitte 2/3 zu Follett und Dewey selektiv gelesen. Zitierte Originalbücher nicht direkt gelesen.

Fragen aus den sieben Kurzbegutachtungen

- **Kant:** Die Untersuchung eigener Gründe und wirksamer Änderungsbeiträge kann erfolgreiche Koordination von gemeinsam begründeter Zwecksetzung abgrenzen.

- **Popper:** Getrennte Beobachtungen von Zielannahme und Vorhabenbeginn können gegensätzliche Wirkungen der Vorleistungsgelegenheit sichtbar machen.

- **Wittgenstein:** Konkrete Grenzfälle zwischen Präzisierung und Aufgabe eines bisherigen Zwecks können die Kriterien neuer gemeinsamer Vorhaben schärfen.

- **Plessner:** Die Rekonstruktion von Vorschlagsänderungen, Einwänden und akzeptierten Ablehnungen kann die erlebte Offenheit gemeinsamer Zweckbildung untersuchen.

- **Cassirer:** Beobachtbare gemeinsame Umdeutungen des Beitrags könnten zeigen, wie neue Zwecke entstehen, ohne einen Vertrauensmechanismus vorwegzunehmen.

- **Friston:** Eine ergänzende Prozessstudie könnte Veränderungen des wahrgenommenen Handlungsrepertoires und ihrer Sicherheit als konkurrierende Erklärungen untersuchen.

- **Luhmann:** Eine getrennte Beobachtung von Zweckinnovation und Anschlusszusagen könnte zeigen, wie Vorleistungen gemeinsame Bindung ermöglichen.

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

Die Einwände Wittgensteins und Fristons nötigen mich, meine erste Stellungnahme an einem bestimmten Punkt einzuschränken: Auch eine ausdrückliche Antwort über Geltung gewährt keinen unmittelbaren Zugang zum vollzogenen Urteilsakt. Meine Unterscheidung zwischen Entbehrlichkeit und Geltungsentzug bleibt notwendig; hinzu tritt jedoch die Unterscheidung zwischen allgemeiner Aberkennung der Verwendbarkeit und einem einer Instanz zugeschriebenen Widerruf. Wittgenstein hat recht, dass erst der bestimmte Gebrauch einer Antwort ihre Zuordnung erlaubt. Das Antwortformat muss deshalb fortbestehende Begründungstauglichkeit, tatsächliche Belegauswahl und angenommene Rücknahme getrennt erfassen, ohne diese Unterscheidungen durch suggestive Fragen selbst hervorzubringen. Mit Cassirer stimme ich darin überein, dass gleiche Angaben keine gleiche symbolische Funktion gewährleisten. Daraus folgt für die Versuchsanordnung eine Grenze: Sie darf die Veränderung der Bedeutung nicht zugleich untersuchen und durch die Forderung vollständiger Bedeutungsgleichheit ausschließen. Poppers direkter Vergleich mit der zeitneutralen Verbindung erhält gerade hier sein Gewicht; sein eigener, vorab begründeter Mindestunterschied verhindert, dass jede beliebige Wirkung der Grafik als Bestätigung gilt. Ich reformuliere daher: Gegenüber sowohl einer unverknüpften Übersicht als auch einer vorvalidierten zeitneutralen Verbindung erhöht die zeitliche Verknüpfung um jeweils vorab begründete Mindestbeträge die regelwidrige Aberkennung fortbestehender Begründungstauglichkeit, wobei der zusätzliche Fehler dem vertauschten Erfassungslabel folgt und ein behaupteter Widerruf gesondert ausgewiesen wird. Diese Fassung schränkt den Gegenstand ein und bekräftigt zugleich seine Prüfbarkeit. Plessners Einwand einer bloß erfüllten Antworterwartung bleibt damit allerdings offen: Ergänzende Begründungen können diese Möglichkeit unterscheiden helfen, ohne sie als inneren Vorgang abschließend nachzuweisen. Die produktive Spannung liegt zwischen dem festzuhaltenden Maßstab richtigen Urteilens und der nur mittelbaren Bestimmung dessen, was tatsächlich geurteilt wurde. Der Maßstab stammt unter den gesetzten Bedingungen aus der Aufgabenregel; welche Regel der Antwortende seiner Antwort zugrunde legt, lässt sich daraus nicht ableiten. Schließlich wäre ein bestätigter Befund eine empirische Fehlanwendung einer Geltungsregel, kein Beweis gegen die Bedingungen vernünftiger Geltungsprüfung.

Replik 2 — Popper

Mit Kant teile ich die Forderung, fortbestehende Geltung und entfallende Entscheidungsrelevanz schon im Antwortformat auseinanderzuhalten. Wittgenstein zwingt mich jedoch, mein erstes Gutachten zu präzisieren: Auch eine ausdrückliche Geltungsaberkennung behauptet noch nicht notwendig einen Widerruf. Deshalb sollten Belegauswahl, weiterhin erlaubte Verwendung zur Begründung derselben Aussage und zugeschriebene Rücknahme getrennt erhoben werden. Für die Hauptthese empfehle ich den unbegründeten Geltungsentzug als primären Gegenstand; ein erfundener Widerruf wäre die enger zu belegende Unterform. Diese Festlegung muss vor der Untersuchung erfolgen und darf keinem günstigeren Antwortmuster folgen. Meine Forderung nach einem vorab bestimmten direkten Kontrast zwischen zeitlicher und zeitneutraler Verbindung bekräftige ich. Entscheidend bleibt dessen Bindung an den randomisierten Labelwechsel, mit eigener begründeter Mindestwirkung und ausreichender Präzision. Cassirers Einwand verändert allerdings die Interpretation dieses Tests: Informationsgleiche Materialien sind nicht bedeutungsgleich. Der Kontrollvergleich isoliert daher keine reine Zeitlichkeit; er prüft die gewählten Darstellungen samt ihren vorvalidierten Konventionen. Mein Reformulierungsbeitrag lautet: Bei vorab validierten Aufgaben erhöht die zeitliche gegenüber sowohl der unverknüpften als auch der zeitneutral verknüpften Darstellung unbegründete Aberkennungen fortbestehender Beleggeltung um vorab begründete Mindestwirkungen, wobei der zusätzliche Fehler dem jeweils früheren Erfassungslabel folgt. Plessners vermutete Prüfungsanforderung und Fristons Auswahlstrategie bleiben ernsthafte Alternativen zur Annahme einer veränderten Geltungsüberzeugung. Getrennte, ausbalancierte Geltungsfragen können diese Alternativen unter Druck setzen; nachgelagerte Selbstauskünfte entscheiden sie nicht. Hier liegt auch meine Grenze gegenüber einer stärkeren Modellinterpretation: Ein passendes Modell ist noch kein ausgeschlossener Konkurrent. Die produktive Spannung bleibt, dass verständliche Zeichen ihre erlernten Erwartungen mitbringen, während der Versuch gerade deren unbegründete Übertragung messen will. Methodisch tragfähig wird dies durch begrenzte Behauptungen und vorab festgelegte Scheiterpunkte. Ein präzise ausgeschlossener spezifischer Kontrast widerlegt die reformulierte Vorhersage im festgelegten Aufgabenbereich; er darf nicht nachträglich durch eine Erzählung über verborgene Autorität gerettet werden.

Replik 3 — Wittgenstein

Kant bestätigt meine erste Unterscheidung: Ein entbehrlicher Beleg hat seine Geltung nicht verloren. Doch Plessner nötigt mich, meinen Vorschlag zu ergänzen. Auch die ausdrückliche Antwort, der frühere Beleg dürfe nicht mehr verwendet werden, kann die vermutete Prüfungsanforderung bedienen. Ein schärferer Fragesatz beseitigt nicht jede Unklarheit seines Gebrauchs. Man muss Beispiele prüfen, in denen dieselbe Person einen Beleg auswählt, seine fortbestehende Verwendbarkeit beurteilt und einen angeblichen Rücknahmeakt beschreibt. Diese Antworten dürfen auseinanderfallen. Gerade dann haben wir etwas gelernt. Cassirer hat recht: Gleiche Daten verbürgen keine gleiche Bedeutung. Die Verbindung kann eine andere Verwendung nahelegen; deshalb sollte „informationsgleich“ nur die verfügbaren Sachangaben bezeichnen. Sein Ausdruck einer symbolischen Leistung wird aber unklar, sobald er die Beschreibung der konkreten Zeichenverwendung ersetzt. Welche Verbindung lesen die Beteiligten als Ablage, welche als Ablösung? Das ist vorab am Material zu prüfen. Fristons Trennung von Geltungsrevision und Auswahlstrategie hilft. Dass er auch solche Urteile nicht als unmittelbare Fenster behandelt, halte ich fest. Den Ausdruck „innerer Widerruf“ würde ich dennoch streichen; er verspricht eine zusätzliche Feststellung, deren Verwendungskriterien erst zu zeigen wären. Popper verlangt zu Recht einen eigenen Scheiterpunkt für den Vergleich mit der zeitneutralen Verbindung. Ohne ihn könnte jede Pfeilwirkung als Wirkung der Chronologie durchgehen. Mein Reformulierungsvorschlag: „Gegenüber unverknüpfter Übersicht und zeitneutraler Verbindung erhöht die zeitliche Verknüpfung unter vorab geprüften Antwortbedingungen die ausdrückliche Aberkennung fortbestehender Beleggeltung; der zusätzliche Fehler folgt dem früheren Erfassungslabel, während bloße Auswahl und zugeschriebene Rücknahme getrennt erfasst werden.“ Damit bekräftige ich meinen ersten Einwand und begrenze seinen Anspruch. Die Fragen müssen die Verwendungen unterscheidbar machen, ohne den untersuchten Gebrauch durch Unterricht erst herzustellen. Diese Spannung bleibt. Nachgelagerte Erläuterungen können sie sichtbar machen; sie dürfen den Primärbefund nicht nachträglich passend sortieren.

Replik 4 — Plessner

Ich halte an der Unterscheidung zwischen der Stellung zum Sachverhalt und der Stellung gegenüber dem Darstellenden fest, muss aber meinen ersten Vorschlag präzisieren. Kant und Wittgenstein zeigen, dass eine nachgelagerte Erläuterung allein nicht genügt: Wenn schon die primäre Antwort Auswahl, Entbehrlichkeit und Geltungsentzug vermischt, bleibt unbestimmt, welches Phänomen gezählt wurde. Die fortbestehende Verwendbarkeit eines Belegs muss deshalb im vorvalidierten Antwortformat von seiner tatsächlichen Auswahl unterscheidbar sein, ohne einen Widerruf sprachlich nahezulegen. Fristons Warnung bekräftigt meinen Vorbehalt: Auch eine ausdrückliche Geltungsantwort eröffnet keinen unmittelbaren Zugang zur Überzeugung des Antwortenden. Sie kann noch Ausdruck einer angenommenen sozialen Anforderung sein. Hier besteht eine Konvergenz, doch die Stellung des Menschen zu seiner eigenen Antwort geht für mich in der Konkurrenz generativer Modelle nicht auf. Mit Luhmann teile ich die Frage nach zugeschriebener Autorität; ich würde jedoch die Erfahrung, sich als Adressat einer sinnvoll gemeinten Darstellung zu verstehen, nicht allein durch die Beschreibung kommunikativer Selektionen ersetzen. Cassirer veranlasst mich überdies, meinen Vorschlag einer automatisch erzeugten Ablagesortierung einzuschränken: Auch Automation kann eine symbolisch vermittelte Verlässlichkeit versprechen und ist daher keine neutrale Entfernung des Darstellenden. Mein Reformulierungsbeitrag lautet: Eine zeitlich verknüpfte Darstellung kann unter festgelegten Aufgabenregeln die ausdrückliche Aberkennung fortbestehender Beleggeltung gegenüber geeigneten Kontrolldarstellungen erhöhen, wobei diese Aberkennung von Auswahl und Entbehrlichkeit getrennt erfasst werden muss und ihre Bindung an das jeweils früher zugewiesene Erfassungslabel noch keine bestimmte Stellungnahme des Urteilenden zu sich und zum Darstellenden erklärt. Poppers geforderter direkter Vergleich mit zeitneutralen Verbindungen bleibt dafür eine notwendige empirische Prüfung; er entscheidet nicht über diese anthropologische Deutung. Getrennte Erläuterungen sollen mögliche Erfahrungsunterschiede erschließen, ohne Personen nachträglich auszuschließen oder den Mechanismus zu beweisen. Die produktive Spannung bleibt bestehen: Wir benötigen öffentlich prüfbare Antworten, obwohl der Mensch sich zu dem, was er antwortet, nochmals verhalten kann.

Replik 5 — Cassirer

Meine erste Deutung eines Wechsels von der Ordnung des Auffindens zur Ordnung der Geltung bleibt bestehen, muss aber an einer entscheidenden Stelle bescheidener werden: Eine symbolische Funktion ist nicht schon dadurch bestimmt, dass der Antwortende das Wort „ersetzt“ verwendet. Kant und Wittgenstein zeigen mit Recht, dass Entbehrlichkeit, Auswahl und Geltungsentzug auseinanderfallen; hinzu kommt die engere Zuschreibung eines Widerrufs durch eine zuständige Instanz. Ich würde deshalb meine vorgeschlagene Erläuterung nicht allein als ergänzende Deutung behandeln, sondern schon die unabhängige Vorprüfung des Antwortformats darauf verpflichten, diese Gebrauchsweisen unterscheidbar zu machen. Die nachgelagerten Erläuterungen bleiben dagegen beschreibend und rechtfertigen weder Ausschlüsse noch Aussagen über innere Vorgänge. Mit Luhmann und Plessner gewinnt meine erste Position eine weitere Bestimmung: Die grafische Form wirkt innerhalb einer Beziehung zu einem vermeintlich verantwortlichen Darstellenden, dessen Ordnungsleistung als sachliche Auskunft verstanden werden kann. Allerdings darf diese Erweiterung die Eigentümlichkeit der Form nicht in zugeschriebener Autorität auflösen; umgekehrt wäre es ebenso voreilig, ihre Wirkung von solchen Zuschreibungen unabhängig zu erklären. Hier bleibt eine produktive Spannung zwischen kulturell erlernter Zeichenbedeutung und situativer Mitteilungsdeutung. Poppers direkter Vergleich mit einer zeitneutralen Verbindung verdient Zustimmung: Die spezifische Chronologiebehauptung muss an ihm scheitern können, obwohl Informationsgleichheit keine Bedeutungsgleichheit herstellt. Mein Reformulierungssatz lautet: Bei gleichem Bestand weiterhin gültiger Belege erhöht eine nach sachlich irrelevanten Erfassungszeiten geordnete Verbindung gegenüber unverknüpfter und zeitneutral verknüpfter Darstellung ausdrückliche, unbegründete Geltungsaberkennungen zulasten des früher etikettierten Belegs, wobei sich dieser zusätzliche Fehler bei unabhängig kontrollierter Darbietungsfolge mit dem Labeltausch verschiebt und eine zugeschriebene Rücknahme gesondert erfasst wird. Darin wird die symbolische Funktion als prüfbare Beziehung untersucht; weder die Chronologie noch die Darstellung erhält den Rang eines selbsttätigen Subjekts.

Replik 6 — Friston

Kant und Wittgenstein präzisieren meinen ersten Einwand: Die Auswahl eines Belegs, die Aberkennung seiner Geltung und die Zuschreibung einer Rücknahme sind verschiedene beobachtbare Antworten; sie dürfen deshalb auch im generative model nicht durch dieselbe Beobachtungsvariable vertreten werden. Ich korrigiere entsprechend meine anfängliche Gegenüberstellung von Geltungsrevision und Auswahlpräferenz: Ein Modell muss zusätzlich unterscheiden, welche Antwort eine Person aufgrund einer vermuteten Prüfungsanforderung berichtet. Plessners Einwand sozialer Erwartungen und Luhmanns Hinweis auf zugeschriebene Revisionsautorität liefern hierfür konkrete Kandidaten, keine bereits nachgewiesenen Mechanismen. Formal beträfe dies sowohl die Priorannahmen über die Funktion der Darstellung als auch das Beobachtungsmodell, das latente Überzeugungen mit Antworten verbindet. Gerade diese doppelte Möglichkeit begrenzt die Identifizierbarkeit: Verschiedene Kombinationen können dieselben Ersetzungsangaben erzeugen. Getrennte, unabhängig vorvalidierte Fragen zur fortbestehenden Begründungsverwendbarkeit, zur Belegauswahl und zu einer ausdrücklich angenommenen Rücknahme verbessern den Vergleich, beseitigen diese Mehrdeutigkeit jedoch nicht automatisch. Poppers Forderung nach einem vorab bestimmten direkten Kontrast zur zeitneutralen Verbindung übernehme ich uneingeschränkt; ohne eigene Mindestwirkung dieses Kontrasts könnte auch ein elegantes Modell die fehlende Chronologiespezifität lediglich umbenennen. Mein Reformulierungsvorschlag lautet: „Bei unverändert gültigen Belegen erhöht zeitliche gegenüber zeitneutraler Verknüpfung die ausdrückliche Aberkennung fortbestehender Beleggeltung um einen vorab begründeten Mindestbetrag; die Richtung dieses zusätzlichen Fehlers folgt dem randomisierten früheren Erfassungslabel.“ Dieser Satz bezeichnet eine prüfbare Wirkung, während der stärkere Ausdruck Widerruf eine zusätzlich erhobene Rücknahmezuschreibung voraussetzt. An meiner ersten Begrenzung halte ich fest: Das freie Energieprinzip allein liefert weder diese Richtung noch die Effektgröße. Die produktive Spannung zu Wittgenstein bleibt, ob ein Vergleich formaler Modelle über unterschiedliche Antwortgebräuche hinaus einen latenten Geltungswechsel identifizieren kann. Dafür wären zusätzliche, zwischen den Modellen diskriminierende Interventionen erforderlich; der Hauptversuch muss diesen Erklärungsanspruch nicht erfüllen.

Replik 7 — Luhmann

Meine erste Stellungnahme ist zunächst an Wittgensteins Einwand zu korrigieren: Die Zurechnung einer Rücknahme durch eine zuständige Instanz darf nicht mit jedem behaupteten Geltungsverlust zusammenfallen. Eine Versuchsperson kann einen Beleg als überholt behandeln, ohne irgendeinen Widerrufenden angeben zu können. Beobachtet werden deshalb verschiedene Kommunikationen: Auswahl, Aberkennung weiterer Verwendbarkeit und Zuschreibung einer Rücknahme. Kant verschärft zu Recht die Grenze zwischen Entbehrlichkeit und Geltungsentzug; der experimentelle Beobachter benötigt dafür jedoch eine vorab validierte Unterscheidung, keinen Zugang zur Urteilskraft des Beobachteten. Plessners Hinweis auf vermutete Prüfungsanforderungen bekräftigt meine These konkurrierender Mitteilungen: Die ausdrückliche Geltungsregel und die grafisch zugeschriebene Aktualisierungsabsicht können unterschiedlich gewichtet werden. Seine anthropologische Begründung übernehme ich dabei nicht; für den Versuch genügt zunächst, wie solche Erwartungen in Antworten kommuniziert werden. Mit Cassirer lässt sich außerdem präzisieren, dass gleiche Beleginformationen keine gleiche Darstellungsbedeutung garantieren. Die Zeichen der zeitneutralen Kontrolle müssen deshalb ebenfalls vorvalidiert werden; sie garantieren nicht die Abwesenheit von Bedeutung. Poppers Forderung nach einem vorab bestimmten direkten Kontrast verhindert, dass dieser unvermeidliche Rest zum allgemeinen Rettungsargument wird. Mein Reformulierungsbeitrag lautet: Unter vorvalidierten Regeln fortbestehender Beleggeltung erhöht eine zeitliche Verknüpfung gegenüber einer unverknüpften Übersicht und einer zeitneutralen Verknüpfung die ausdrückliche, regelwidrige Aberkennung weiterer Begründungsverwendbarkeit des jeweils früher etikettierten Belegs, wobei der zusätzliche Fehler unabhängig von der tatsächlichen Darbietungsfolge dem randomisierten Etikettenwechsel folgt. Damit wird ein engerer Gegenstand geprüft, als der Titel vom erfundenen Widerruf suggeriert. Meine vorgeschlagene Variation zwischen Ablageverzeichnis und redaktionellem Geltungsstand bleibt eine Anschlussstudie zur Autoritätszurechnung. Ungelöst bleibt, ob eine erklärte Aberkennung eine Geltungsannahme oder die Anpassung an eine vermutete Antworterwartung bezeichnet. Auch zusätzliche Erläuterungen sind Kommunikationen; sie beobachten diese Spannung weiter, heben sie aber nicht auf.

Sokrates-Synthese

Die folgende Stimme ist eine KI-Rekonstruktion in einem eigenen GPT-6-Astra/max-Kontext. Sie erhielt die vollständigen sieben Hauptgutachten und sieben Repliken, nicht die späteren externen Prüfungen.

Bewegung 1 — Konvergenzen

Wenn ich die sieben Hauptgutachten und sieben Repliken als KI-rekonstruierte Perspektiven recht verstehe, stimmen Kant, Popper und Wittgenstein in einer entscheidenden Unterscheidung überein: Einen Beleg wegzulassen bedeutet weder, ihm die Geltung abzusprechen, noch, jemandem seinen Widerruf zuzuschreiben. Gerade Kant und Wittgenstein verlangen deshalb, fortbestehende Begründungsverwendbarkeit und entfallende Entscheidungsrelevanz auseinanderzuhalten. In den Repliken übernehmen auch Plessner, Cassirer, Friston und Luhmann die Forderung, diese Unterschiede bereits im unabhängig vorgeprüften Antwortformat kenntlich zu machen. Nachträgliche Erläuterungen können eine unklare primäre Messung nicht heilen. Poppers zweite gemeinsame Forderung betrifft den direkten Vergleich mit einer zeitneutralen Verbindung: Auch dieser Vergleich muss einen eigenen Scheiterpunkt besitzen und den zusätzlichen Fehler an den randomisierten Labeltausch binden. Cassirer schärft dabei die Reichweite: Gleiche Sachangaben gewährleisten keine gleiche Bedeutung. Worauf können wir uns somit verpflichten? Auf einen begrenzten Darstellungsbefund mit getrennten Antwortmaßen, vorab begründeten Mindestwirkungen und einem ernsthaften Kontrollvergleich. Ob die Antworten einen inneren Geltungswechsel anzeigen, bleibt selbst bei einem passenden Ergebnis offen; darin treffen sich besonders Plessners und Fristons Vorbehalte.

Bewegung 2 — Divergenzen

Wo widersprechen sich die Stimmen tatsächlich? Begrifflich korrigiert Wittgenstein die zunächst zu weite Verwendung von Widerruf: Eine Aberkennung weiterer Verwendbarkeit braucht noch keinen behaupteten Rücknahmeakt. Kant und Luhmann nehmen diese Korrektur auf; hier bleibt keine unaufgelöste Gegenthese zurück. Methodisch besteht dagegen Spannung zwischen Fristons möglichen Modellvergleichen und Wittgensteins Forderung nach öffentlichen Verwendungskriterien. Friston hält eine weitergehende Identifikation innerer Annahmen durch zusätzliche, unterscheidende Eingriffe für möglich; Wittgenstein bestreitet, dass der Ausdruck eines inneren Widerrufs schon hinreichend bestimmte Kriterien besitzt. Substanziell verweigert Plessner die Reduktion des Sichverhaltens zur eigenen Antwort auf Modelle oder kommunikative Auswahl, während Luhmann für den Versuch bei kommunizierten Erwartungen ansetzt. Cassirer wiederum lässt die Eigenart der Zeichen weder in zugeschriebener Autorität aufgehen noch erklärt er sie vorschnell für autoritätsunabhängig. Diese Unterschiede liefern keine direkt entgegengesetzten Vorhersagen für den engen Hauptversuch. Welche Beobachtung würde sie entscheiden? Poppers Antwort begrenzt zunächst den Anspruch: Der Kontrollvergleich kann eine Darstellungsbehauptung prüfen, aber keine dieser umfassenderen Erklärungen allein bestätigen. Der verbleibende Streit betrifft folglich vor allem den erklärenden Schluss aus einem möglichen Befund.

Bewegung 3 — Produktive Antinomien

Drei Spannungen sollten wir festhalten, ohne sie zu vermeintlichen Lösungen zu verkürzen. Erstens verlangt Kant einen unabhängig bestimmten Maßstab richtiger Belegverwendung, während Plessner und Friston daran erinnern, dass die beobachtete Antwort diesen Maßstab verfehlen kann, ohne eine entsprechend veränderte Überzeugung unmittelbar offenzulegen. Zweitens braucht Poppers Kontrollvergleich unterscheidbare, vorgeprüfte Darstellungen; mit Cassirer bleibt jedoch bestehen, dass verständliche Zeichen ihre erlernten Bedeutungen mitbringen. Vollständige Bedeutungsgleichheit würde gerade die untersuchte Veränderung beseitigen, während unkontrollierte Bedeutungsunterschiede ihre Zuschreibung erschweren. Drittens hält Cassirer an der besonderen Leistung der grafischen Form fest, während Luhmann und Plessner die zugeschriebene Absicht oder Autorität des Darstellenden hervorheben. Form und Zuschreibung können zusammenwirken, ohne deshalb dieselbe Erklärung zu sein. Welche Arbeit leisten diese Antinomien? Sie begrenzen, was aus einem möglichen Effekt folgen darf, und verlangen unterscheidbare Anschlussfragen. Es handelt sich hier um produktive Erkenntnis- und Prüfspannungen, nicht um logische Unwiderlegbarkeit. Keine davon darf nach einem präzise gescheiterten Kontrollvergleich als Rettung der Hauptvorhersage dienen; ebenso wenig darf ein erfolgreicher Vergleich sie als erledigt ausgeben.

Bewegung 4 — Reformulierungs-Anstoß

Welcher Vorschlag trägt diese Einwände am weitesten? Poppers Replik verbindet die ausdrückliche Geltungsaberkennung mit beiden Kontrollvergleichen und verlangt für den zeitneutralen Vergleich eine eigene Mindestwirkung. Wittgensteins Unterscheidung der Antworten und Cassirers Bindung an den kontrollierten Labelwechsel machen daraus eine präzisere Fassung. Der primäre Gegenstand wird dabei ausdrücklich geändert: Geprüft wird die allgemeinere Aberkennung fortbestehender Begründungsverwendbarkeit; ein zugeschriebener Widerruf ist ihre enger zu belegende Unterform und wird separat erhoben. Das erweitert den erfassten Gegenstand und schwächt den Erklärungsanspruch gegenüber dem ursprünglichen Titel. Es bestätigt die stärkere Ausgangsfassung nicht. Die Intervention bleibt dieselbe zeitliche Verknüpfung nach sachlich irrelevanten Erfassungszeiten. Auch sachliche Gültigkeit, unabhängige Belege, unveränderter Sachverhalt und kontrollierte Darbietungsfolge bleiben Voraussetzungen. Der zusätzliche Fehler muss gegenüber beiden Vergleichsdarstellungen mit dem früheren Label wandern; ein allgemeiner Labeleffekt genügt nicht. Mindestwirkungen und Präzision sind noch zu begründen, keine vorhandenen Befunde. Luhmanns Autoritätsvariation bleibt eine Anschlussstudie, damit wir die offene Erklärung nicht unbemerkt in eine andere Hauptintervention verwandeln.

Bewegung 5 — Offene Frage

Es bleibt zu fragen: Welche zusätzliche Intervention liefert unterschiedliche Vorhersagen für eine veränderte Geltungsannahme und die Anpassung an eine vermutete Antworterwartung, wenn beide dieselbe ausdrückliche Geltungsaberkennung erzeugen können? Diese Frage erhält den Tag #verzweigung-offen-geltungsaberkennung-antworterwartung. Friston verlangt hierfür vorab unterscheidbare Modelle und Eingriffe; Wittgenstein verlangt zugleich, dass deren Unterschiede im konkreten Antwortgebrauch erkennbar werden. Luhmanns Variation zwischen Ablageverzeichnis und redaktionellem Geltungsstand bietet einen möglichen Ansatz, identifiziert aber für sich noch keine innere Überzeugung. Plessners Einschränkung ist ebenfalls festzuhalten: Eine automatisch erzeugte Darstellung entfernt zugeschriebene Verlässlichkeit nicht einfach. Auch Cassirers Zeichenkonventionen müssen deshalb in einer solchen Anschlussstudie berücksichtigt werden. Produktiv wird der Anschluss, sobald unabhängige Vorprüfungen die Antwortunterscheidungen und beide Verbindungskontrollen tragen und für die neue Intervention konkurrierende Vorhersagen formuliert sind. Nachgelagerte Selbstauskünfte können dafür Material liefern, aber keine Entscheidung erzwingen. Der Hauptversuch bleibt unterdessen eigenständig prüfbar; sein Scheitern dürfte nicht durch die Verlagerung auf diese offene Frage verdeckt werden. So kann der Chronologie-Strang fortgesetzt werden, ohne aus einem beobachtbaren Antwortmuster vorschnell einen verborgenen Vorgang zu machen.

Finale Hypothese

Zeitliche Verknüpfung begünstigt unbegründete Geltungsaberkennung

Kernsatz. Bei weiterhin gültigen, sachlich unabhängigen Belegen zum selben unveränderten Sachverhalt erhöht eine grafische Verknüpfung nach sachlich irrelevanten Erfassungszeiten die ausdrückliche Aberkennung ihrer weiteren Verwendbarkeit als Begründung gegenüber unverknüpften und zeitneutral verknüpften Darstellungen mit denselben Sachangaben. Der zusätzliche Fehler trifft den jeweils früher etikettierten Beleg und folgt dem Tausch der Erfassungslabels, auch wenn die tatsächliche Darbietungsfolge und die Expositionszeiten unabhängig kontrolliert werden.

Strukturthese. nicht separat beansprucht

Empiriethese. Für den vorab festgelegten und unabhängig validierten Aufgabenbereich werden gegenüber sowohl unverknüpften als auch zeitneutral verknüpften Darstellungen praktisch relevante zusätzliche ausdrückliche Aberkennungen fortbestehender Begründungsverwendbarkeit vorausgesagt;

der jeweilige zusätzliche Fehler muss der randomisierten Zuordnung des früheren Erfassungslabels folgen. Die hierfür erforderlichen Mindestwirkungen sind vor der Hauptstudie eigenständig zu begründen und festzulegen.

Begründung. Die Hypothese setzt den in 06 Hypothesentag/2026-10-19 dokumentierten Chronologie-Strang fort, indem sie den möglichen Übergang von einer Ablageordnung zu einer unbegründeten Bewertung gültiger Belege prüft. Gegenüber dem ursprünglichen Titel „Die Chronologie erfindet einen Widerruf“ wird der primäre Gegenstand ausdrücklich zur allgemeineren Geltungsaberkennung erweitert; ein einer Instanz zugeschriebener Widerruf wird gesondert erhoben und durch einen Befund zur Geltungsaberkennung nicht bestätigt. Als Fehler zählt die ausdrückliche Behauptung, der Beleg dürfe dieselbe Aussage nicht mehr begründen, obwohl die Aufgabenregel diese Verwendung weiterhin erlaubt; das Verständnis dieser Regel und die Unterscheidbarkeit der Antworten müssen vor der Hauptstudie unabhängig validiert werden. Bloße Auswahl, Entbehrlichkeit und angenommene Rücknahme werden getrennt erfasst, damit unterschiedliche Antwortverwendungen nicht gleichgesetzt werden. Die Vorhersage betrifft die untersuchten Darstellungen und beobachtbaren Antworten; weder ein innerer Überzeugungswechsel noch ein bestimmter Aufmerksamkeits-, Bedeutungs- oder Autoritätsmechanismus wird damit bereits nachgewiesen.

Falsifikationsbedingung. Im vorab festgelegten Aufgabenbereich scheitert die auf praktisch relevante Mindestwirkungen präzisierte Vorhersage bei gültiger Durchführung, sobald die obere Grenze eines nach vorab bestimmter Fehlerkontrolle berechneten Unsicherheitsintervalls für mindestens einen der beiden Darstellungsvergleiche oder dessen gerichtete Bindung an den Labeltausch unter der jeweils begründeten Mindestwirkung liegt. Breite Intervalle und fehlende statistische Signifikanz entscheiden den Test nicht; solange Materialvalidierung, Mindestwirkungen und Präzisionsplanung fehlen, ist seine Durchführung noch nicht vorbereitet.

Prüfentwurf

Voraussetzungen: Vor der Hauptuntersuchung werden Aufgaben mit sachlich unabhängigen, weiterhin gültigen Belegen zu demselben unveränderten Sachverhalt sowie deren Antwortformate an unabhängigem Material und einer gesonderten Stichprobe geprüft. Erfassungszeit bedeutet ausschließlich Ablagezeit, keine Messzeit, Revision oder Qualitätsstufe; die Aufgabenregel erlaubt die weitere Verwendung beider Belege zur Begründung derselben Aussage. Unterschiedlicher sachlicher Informationswert wird kontrolliert, weil Quellenunabhängigkeit allein nicht genügt. Die Vorprüfung muss nach vorab bestimmten Kriterien zeigen, dass Auswahl, Entbehrlichkeit, ausdrückliche Aberkennung weiterer Begründungsverwendbarkeit und zugeschriebene Rücknahme unterscheidbar erhoben werden können, ohne die vermutete Fehlerquelle durch suggestive Fragen oder Unterricht vorzugeben. Diese Validierung steht noch aus. Unabhängige Faktoren: Randomisiert werden die Darstellungsform mit drei Bedingungen T, U und N, die Zuordnung früherer und späterer Erfassungslabels zu inhaltlich festgehaltenen Belegen sowie die tatsächliche Reihenfolge ihrer Erstexposition. T verbindet die Belege grafisch nach Erfassungszeit, U zeigt dieselben Sachangaben und Zeitlabels unverknüpft, N verwendet eine möglichst geometrisch entsprechende, vorab geprüfte Verbindung ohne zeitliche Bedeutung. Position, Hervorhebung, Lesbarkeit und Informationsmenge werden abgeglichen oder unabhängig ausbalanciert; Dauer und Abstände der tatsächlichen Exposition werden kontrolliert. Danach sind sämtliche Belege in allen Bedingungen gleich lange gleichzeitig verfügbar. Gleiche Sachangaben bedeuten keine vollständige Bedeutungsgleichheit; verbleibende Unterschiede der Zeichenkonventionen werden dokumentiert. Primäre Endpunkte: Für jeden Beleg wird regelbasiert erfasst, ob seine weiterhin erlaubte Verwendung als Begründung ausdrücklich aberkannt wird. Bloße Nichtauswahl und fehlender zusätzlicher Nutzen zählen nicht als Aberkennung. Getrennte Fragen erheben Belegauswahl und die Zuschreibung eines Rücknahmeakts; Antwortformulierungen und Antwortpositionen werden unabhängig vorgeprüft und ausbalanciert. Ein behaupteter Widerruf erfordert eine zusätzliche Rücknahmezuschreibung und wird nicht aus dem primären Fehler abgeleitet. R_D bezeichnet in Darstellungsbedingung D den Anteil ausdrücklicher Aberkennungen über beide Belege und beide Labelzuweisungen. B_D bezeichnet die Differenz der Aberkennungswahrscheinlichkeiten für einen früher gegenüber einem später etikettierten Beleg, über die inhaltlich festgehaltenen Belege, beide randomisierten Labelzuweisungen und tatsächlichen Darbietungsfolgen gemittelt. Für jeden Kontrollvergleich C aus U und N werden sowohl der zusätzliche Fehler R_T minus R_C als auch dessen gerichtete Labelbindung B_T minus B_C geschätzt. Die zweite Größe ist die Wechselwirkung von Darstellungsform und Labelzuordnung; ein in allen Darstellungen gleicher Labeleffekt erfüllt diese Vorhersage nicht. Entscheidungsschwellen: Für alle vier notwendigen Kontraste werden vor der Hauptstudie positive praktisch relevante Mindestwirkungen sachlich begründet und festgelegt, insbesondere eigenständige Grenzen für T gegenüber N. Es werden hier weder Zahlen eingesetzt noch vorhandene empirische Begründungen behauptet. Stichprobengröße, Modell, Fehlerkontrolle und Intervallverfahren werden auf diese Konjunktion und auf die Abhängigkeiten wiederholter Urteile innerhalb von Personen und Aufgaben ausgelegt. Unterstützung der präzisierten Vorhersage verlangt, dass die nach diesem Verfahren bestimmten unteren Intervallgrenzen die jeweiligen Mindestwirkungen erreichen oder überschreiten; liegt eine obere Intervallgrenze darunter, scheitert mindestens ein notwendiger Bestandteil. Alle übrigen Intervallkonstellationen bleiben für die Konjunktion unentschieden. Sehr kleine positive Effekte können damit die praktische Mindestvorhersage verfehlen, ohne dass jeglicher Effekt ausgeschlossen wäre. Durchführung und Reichweite: Kriterien ungültiger Durchführung werden vorab auf Studienebene festgelegt; alle zugewiesenen Personen bleiben in der primären Analyse, insbesondere unabhängig von nachgelagerten Verständnisantworten. Der Umgang mit fehlenden Antworten wird vorab bestimmt. Absolute Kontraste, Unsicherheit und Sensitivität gegenüber tatsächlicher Expositionsfolge werden berichtet. Nachgelagerte Erläuterungen dienen ausschließlich ergänzender Beschreibung, weder der nachträglichen Fallauswahl noch dem Mechanismusnachweis. Ein von der Grafik veränderter Lese- oder Aufmerksamkeitsverlauf kann zum Verfahrenseffekt gehören. Eine

Variation zugeschriebener Autorität und ein unterscheidender Vergleich innerer Modelle gehören in gesonderte Anschlussstudien; der Hauptversuch behauptet keine unabhängig davon identifizierte Wirkung reiner Zeitlichkeit und gilt nicht für echte Rücknahmen oder veränderliche Sachverhalte.

Begriffliche Änderung gegenüber dem Ausgangspunkt

Der bisherige Titel und Teile der Erstbewertung deuteten falsche Ersetzungsurteile als Zuschreibung eines nicht erfolgten Widerrufs. Nach den Gutachten wird vor der Untersuchung ausdrücklich die allgemeinere, regelwidrige Aberkennung fortbestehender Begründungsverwendbarkeit zum primären Gegenstand. Das erweitert die Extension der Zielgröße und begrenzt zugleich den Erklärungsanspruch: Ein Widerruf verlangt zusätzlich die Zuschreibung eines Rücknahmeakts und wird gesondert erhoben. Ein Befund zur allgemeineren Geltungsaberkennung bestätigt deshalb weder den engeren Widerrufsanspruch noch einen inneren Überzeugungswechsel. Die grafische Intervention nach sachlich irrelevanten Erfassungszeiten bleibt erhalten; die begriffliche Änderung ist keine empirische Bestätigung der Ausgangsfassung.

Reichweite und ungelöste Spannungen

these — Die Fassung behauptet einen begrenzten Darstellungseffekt mit unabhängig variierbaren Faktoren und konkreten Widerlegungsbedingungen; sie führt den dokumentierten Chronologie-Strang fort, ohne bereits einen inneren Mechanismus oder die Architektur eines Forschungsprogramms zu beanspruchen.

- Unabhängig bestimmte Begründungsgeltung versus nur mittelbar deutbare Antwort: Kant verlangt einen festen Fehlermaßstab, Plessner und Friston begrenzen den Schluss von der Antwort auf eine innere Geltungsannahme.

- Kontrollierbare Darstellung versus unvermeidliche Bedeutungskonvention: Poppers entscheidbarer Vergleich benötigt vorgeprüfte Zeichen, deren Bedeutungsunterschiede nach Cassirer nicht vollständig beseitigt werden können, ohne die untersuchte Veränderung zu entfernen.

- Eigenart der grafischen Form versus zugeschriebene Mitteilungsabsicht und Autorität: Cassirer, Luhmann und Plessner eröffnen verschiedene, möglicherweise zusammenwirkende Erklärungen, die der Hauptversuch nicht identifiziert.

Interne Neubewertung

Kriterium	Score	Begründung
Originalität	6	Die Kombination aus irrelevanter Erfassungszeit, Labeltausch und zwei Darstellungskontrollen grenzt einen eigenständigen Prüfgegenstand ab, doch die Neuheit der nun allgemeineren Geltungsaberkennung bleibt ohne weitergehende Literaturprüfung offen.
Falsifizierbarkeit	8	Beide Darstellungsvergleiche und ihre gerichteten Labelbindungen besitzen explizite Scheiterbedingungen, während Antwortvalidierung, sachlich begründete Mindestwirkungen und Präzisionsplanung noch fehlen.
Begriffliche Klarheit	8	Auswahl, Entbehrlichkeit, ausdrückliche Geltungsaberkennung und zugeschriebener Widerruf sind getrennt definiert, ihre zuverlässige Unterscheidung im tatsächlichen Antwortgebrauch ist jedoch noch nicht validiert.
Tiefe	7	Der mögliche Übergang von Ablageordnung zu Begründungsgeltung berührt eine epistemische Grundunterscheidung, ohne einen inneren Urteilsvorgang oder eine allgemeine Theorie der Zeitlichkeit nachzuweisen.
Forschungsrelevanz	8	Die im gelieferten Material dokumentierten Bezüge zu Erklärungsaktualisierung und fortwirkender Fehlinformation tragen den Forschungsanschluss, liefern aber keinen Befund für die hier behauptete Wirkung irrelevanter Erfassungszeiten.
Interdisziplinäre Anschlussfähigkeit	8	Der Entwurf verbindet kognitive Psychologie, Informationsvisualisierung, Mensch-Computer-Interaktion und Erkenntnistheorie, während weitergehende anthropologische und systemtheoretische Mechanismen Anschlussfragen bleiben.
Vault-Anschluss	6	Die Fortsetzung von 06 Hypothesentag/2026-10-19 ist nachvollziehbar und der gelieferte Anschluss an Luhmann - Vertrauen als Komplexitätsreduktion verwendbar, ohne dass damit mehrere vertiefte Sachknoten nachgewiesen wären.
Antinomie-Test	7	Die plausible Gegenthese einer bloßen Auswahl-, allgemeinen Verknüpfungs- oder Expositionswirkung bleibt prüfbar, während die drei festgehaltenen Deutungsspannungen den zulässigen Höchstwert des empirischen Methodentyps nicht erhöhen.
Publikationsmöglichkeit	7	Ein begrenzter experimenteller Beitrag im bereits genannten Umfeld von Memory & Cognition ist vorstellbar, doch Materialvalidierung, begründete Testgrenzen, Daten und eine vertiefte Neuheitsprüfung stehen aus.
Summe /90	**64**	Ausgangswert 68; Änderung -4.

Die Erstbewertung ergibt mit den tatsächlichen Gewichten 1,6 / 1,0 / 1,0 / 1,2 / 1,0 / 1,0 / 1,0 / 1,2 / 1,0 eine Rohsumme von 75,0; bei Gewichtssumme 10,0 und Normierung 0,9 sind 67,5 kaufmännisch auf 68 zu runden. Die finale Rohsumme beträgt 71,4; 71,4 × 0,9 = 64,26, kaufmännisch gerundet 64, somit Differenz −4. Originalität sinkt von 7 auf 6, weil die Neuheit des allgemeineren Gegenstands ungesichert bleibt; Falsifizierbarkeit und begriffliche Klarheit sinken jeweils von 9 auf 8, weil der präzisere Entwurf die noch fehlende unabhängige Validierung und Festlegung der Testgrenzen deutlicher sichtbar macht. Die übrigen Scores bleiben unverändert, insbesondere Antinomie-Test 7 als Obergrenze des empirischen Methodentyps. Inhaltlich verbessert wurden die Trennung der Zielgrößen, die explizite Dokumentation der Begriffsänderung und der eigenständige Scheiterpunkt gegenüber der zeitneutralen Verbindung. Die niedrigere Summe ist kein Nachweis eines gescheiterten Lernprozesses: Die Erstbewertung war

hinsichtlich der bereits erreichten Messklarheit und Prüfreife zu großzügig, und die Schlussfassung erhebt einen bescheideneren Anspruch. Die Scores vergleichen zudem zwei ausdrücklich unterschiedlich abgegrenzte Fassungen; sie messen weder empirische Bestätigung noch den Erfolg eines Experiments.

Offene Anschlussfrage

#verzweigung-offen-geltungsaberkennung-antworterwartung — Welche zusätzliche Intervention liefert unterschiedliche Vorhersagen für eine veränderte Geltungsannahme und die Anpassung an eine vermutete Antworterwartung, wenn beide dieselbe ausdrückliche Geltungsaberkennung erzeugen können?

Pickup-Anlass. Nach unabhängiger Validierung der getrennten Antwortmaße und beider Verbindungskontrollen aufgreifen, sobald für eine zusätzliche Variation zugeschriebener Darstellungsfunktion oder Autorität vorab unterscheidbare Vorhersagen formuliert sind; die Anschlussstudie darf einen gescheiterten Hauptvergleich nicht ersetzen.

Bewertung nach der externen Prüfung

Die interne Bewertung bleibt als eigener Verfahrensstand erhalten. Die folgende Neufassung berücksichtigt Literaturabgleich, frischen Falsifikationsauftrag und schulferne Kritik.

Kriterium	Intern	Nach externer Prüfung	Begründung
Originalität	6	5	Oh et al. (2026) publizieren bereits den Tausch von Erfassungsdaten bei festen Texten; Heiser und Tversky sowie Norris et al. begrenzen zusätzlich den allgemeinen Neuheitsanspruch, während die konkrete Kombination des menschlichen Grafik- und Gültigkeitstests offen bleibt.
Falsifizierbarkeit	8	7	Vier notwendige Kontraste sind festgelegt, aber gemeinsam mögliche Mindestwirkungen, nicht erfolgsabhängige Materialauswahl und die gemeinsame Präzisionsplanung müssen erst unabhängig begründet werden.
Begriffliche Klarheit	8	8	Auswahl, Entbehrlichkeit, Geltungsaberkennung und Rücknahme sind begrifflich getrennt; die Hacking-Prüfung fordert zusätzlich den noch ausstehenden Nachweis vergleichbarer Antwortverwendung unter allen drei Darstellungen.
Tiefe	7	7	Der Übergang von Ablageordnung zu Begründungsverwendbarkeit bleibt eine gehaltvolle Frage, deren Aussage zunächst auf eine eingerichtete Aufgabenpraxis beschränkt ist.
Forschungsrelevanz	8	8	Aktuelle Forschung zu zeitlichen Metadaten und etablierte Arbeiten zu Diagrammen und Informationsprüfung bieten konkrete Vergleichspunkte, ohne die menschliche Hauptvorhersage empirisch zu bestätigen.
Interdisziplinäre Anschlussfähigkeit	8	8	Psychologie, Informationsvisualisierung, Mensch-Computer-Interaktion und Erkenntnistheorie können an den abgegrenzten Test anschließen; die institutionelle Vergleichsstudie bleibt ein eigenes Vorhaben.
Vault-Anschluss	6	6	Der dokumentierte Chronologie-Strang wird mit einem konkreten Material- und Messproblem fortgesetzt; die Quellenarbeit ergänzt ihn, ersetzt aber keine zusätzliche vertiefte Anbindung an mehrere bestehende Sachknoten.
Antinomie-Test	7	7	Allgemeine Verknüpfungswirkung, Antwortkonvention und tatsächliche Expositionsfolge bleiben ernsthafte Alternativen; die empirische Typobergrenze 7 wird eingehalten.
Publikationsmöglichkeit	7	6	Der mögliche Beitrag muss sich gegenüber nahen Vorarbeiten durch einen validierten menschlichen Vergleich ausweisen; eine akademische Publikation benötigt Materialpilot, begründete Grenzen und Daten.
Summe /90	**64**	**61**	Keine empirische Bestätigung.

Die externe Bewertung beträgt 61/90 (Rohsumme 67.8; Faktor 0,9), gegenüber intern 64 und initial 68. Originalität sinkt 6→5 wegen naher publizierter Metadaten- und Zeitstempelarbeiten; Falsifizierbarkeit 8→7 wegen noch nicht verbindlich begründeter, gemeinsam zu planender Mindestwirkungen; Publikationsmöglichkeit 7→6 wegen des engeren zusätzlichen Beitrags und fehlender Daten. Begriffliche Klarheit bleibt 8: Die Definitionen sind getrennt, aber ihre vergleichbare Verwendung muss unter sämtlichen Darstellungen vorab validiert werden. Ein präziserer Entwurf ist damit kein bestätigter oder fertiger Versuch. Die Einwände ändern weder rückwirkend die internen Gutachten noch den gewählten Hauptlinienprozess; die Scores der nur kurz geprüften Nebenlinie sind mit dieser vertieften externen

Bewertung nicht gleichzusetzen.

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

Empirie-Score. 6/10 — Beobachtbare Konsequenzen und Prüfpfad sind ausgearbeitet; validiertes Material, begründete Mindestwirkungen und eigene Daten fehlen.

Empirische Konsequenzen

1. **Zusätzliche Aberkennung gegenüber der Übersicht.** Die zeitliche Verbindung erhöht die ausdrückliche, aufgabenwidrige Aberkennung weiterhin zulässiger Begründungen gegenüber denselben unverknüpften Informationen. Beobachtbar in: randomisierten Urteilsaufgaben mit getrennten Maßen für Belegauswahl und Gültigkeit.

2. **Eigenständiger Effekt gegenüber neutraler Verbindung.** Der Mehrfehler bleibt im direkten Vergleich zu einer möglichst ähnlich gestalteten, aber zeitneutralen Verbindung bestehen. Beobachtbar in: vorgeprüften Grafikpaaren mit gleichen Inhalten, Positionen und Expositionsbedingungen.

3. **Gerichtete Bindung an Erfassungslabels.** Der zusätzliche Fehler wechselt mit der zufälligen Zuweisung des früheren Labels zwischen den inhaltlich festgehaltenen Belegen; ein in allen Darstellungen gleicher Labeleffekt genügt nicht. Beobachtbar in: den Wechselwirkungen von Darstellungsform und Labelzuweisung bei unabhängig ausbalancierter Erstexposition.

Bestehende Befunde

Zu Konsequenz 1

Stand: offen; benachbarte Befunde gemischt. Ecker, Lewandowsky, Cheung und Maybery (2015), He did it! She did it! No, she did not! unterscheiden tatsächliche Aktualität von Präsentationsposition und Rücknahme. Ohne Rücknahme wurde keine signifikante Bevorzugung der zweiten Erklärung beobachtet; dies ist kein starker Nullbeweis, begrenzt aber die Erwartung einer allgemeinen automatischen Ersetzung. Hey, Schaper und Bayen (2026), The continued influence effect and situation models behandeln tatsächliche Widerrufe und gemessene Situationsmodelle, nicht unbegründete Geltungsaberkennung. Norris, Voida, Palen und Voida (2019), Is the Time Right Now? dokumentieren im Feld Konflikte zwischen Zeitstempeln und sachlicher Einordnung; der kausale Laborvergleich bleibt offen.

Zu Konsequenz 2

Stand: offen. Heiser und Tversky (2006), Arrows in Comprehending and Producing Mechanical Diagrams zeigen eine bekannte bedeutungsverändernde Funktion von Pfeilen: Diagramme werden eher dynamisch beziehungsweise funktional verstanden. Das motiviert die neutrale Verbindung als ernsthaften Kontrollvergleich, sagt aber keinen zusätzlichen Fehler allein durch zeitliche Verknüpfung voraus. Min (2023; online 2022), Because it is brand new! findet unterschiedliche Wirkungen von Aktualitätsangaben je nach Produktkontext; ein nicht signifikanter Effekt bei Selbsthilfebüchern ist weder Äquivalenznachweis noch Beleg für die heutigen Materialien. Kontextfreie Aktualitätspräferenz wäre eine zu grobe Begründung.

Zu Konsequenz 3

Stand: offen für Menschen und diese Zielgröße; ein verwandtes Metadatenexperiment liegt vor. Oh, Jwa, Kim und Yoon (2026), Collection-Date Assignment Redirects Answers in Temporal QA tauschen Erfassungsdaten zwischen festgehaltenen Revisionstexten. Antworten mehrerer Sprachmodelle ändern sich; die Wirkung variiert mit Modell und Datumsnähe. Der Preprint vom 23. Oktober nimmt damit die allgemeine Labeltausch-Logik vorweg, untersucht aber konkurrierende historische Sachstände und Antwortauswahl ohne unseren menschlichen Grafikvergleich. Patrick, Bott, Morgan und King (2012), Xu und Kwok (2019) sowie Rottman und Zhang (2025) liefern benachbarte Reihenfolge-, Erinnerungs- und Kausallernbefunde. Keiner dieser gelesenen Befunde ersetzt den Nachweis einer zusätzlichen gerichteten Geltungsaberkennung bei unveränderter Sachlage.

Riskante Vorhersage (Schwellentest)

Vorhersage. Im unabhängig validierten Aufgabenbereich führt T sowohl gegenüber U als auch gegenüber N zu praktisch relevanten zusätzlichen ausdrücklichen Geltungsaberkennungen; in beiden Kontrasten richtet sich der zusätzliche Fehler nach dem randomisierten früheren Erfassungslabel. Es ist eine Konjunktion von vier notwendigen, separat zu prüfenden Mindestwirkungen.

Methodenvorschlag. Voraussetzungen: Vor der Hauptuntersuchung werden Aufgaben mit sachlich unabhängigen, weiterhin gültigen Belegen zu demselben unveränderten Sachverhalt sowie deren Antwortformate an unabhängigem Material und einer gesonderten Stichprobe geprüft. Erfassungszeit bedeutet ausschließlich Ablagezeit, keine Messzeit, Revision oder Qualitätsstufe; die Aufgabenregel erlaubt die weitere Verwendung beider Belege zur Begründung derselben Aussage. Unterschiedlicher sachlicher Informationswert wird kontrolliert, weil Quellenunabhängigkeit allein nicht genügt. Die Vorprüfung muss nach vorab bestimmten Kriterien zeigen, dass Auswahl, Entbehrlichkeit, ausdrückliche Aberkennung weiterer Begründungsverwendbarkeit und zugeschriebene Rücknahme unterscheidbar erhoben werden können, ohne die vermutete Fehlerquelle durch suggestive Fragen oder Unterricht vorzugeben. Diese Validierung steht noch aus. Unabhängige Faktoren: Randomisiert werden die Darstellungsform mit drei Bedingungen T, U und N, die Zuordnung früherer und späterer Erfassungslabels zu inhaltlich festgehaltenen Belegen sowie die tatsächliche Reihenfolge ihrer Erstexposition. T verbindet die Belege grafisch nach Erfassungszeit, U zeigt dieselben Sachangaben und Zeitlabels unverknüpft, N verwendet eine möglichst geometrisch entsprechende, vorab geprüfte Verbindung ohne zeitliche Bedeutung. Position, Hervorhebung, Lesbarkeit und Informationsmenge werden abgeglichen oder unabhängig ausbalanciert; Dauer und Abstände der tatsächlichen Exposition werden kontrolliert. Danach sind sämtliche Belege in allen Bedingungen gleich lange gleichzeitig verfügbar. Gleiche Sachangaben bedeuten keine vollständige Bedeutungsgleichheit; verbleibende Unterschiede der Zeichenkonventionen werden dokumentiert. Primäre Endpunkte: Für jeden Beleg wird regelbasiert erfasst, ob seine weiterhin erlaubte Verwendung als Begründung ausdrücklich aberkannt wird. Bloße Nichtauswahl und fehlender zusätzlicher Nutzen zählen nicht als Aberkennung. Getrennte Fragen erheben Belegauswahl und die Zuschreibung eines Rücknahmeakts; Antwortformulierungen und Antwortpositionen werden unabhängig vorgeprüft und ausbalanciert. Ein behaupteter Widerruf erfordert eine zusätzliche Rücknahmezuschreibung und wird nicht aus dem primären Fehler abgeleitet. R_D bezeichnet in Darstellungsbedingung D den Anteil ausdrücklicher Aberkennungen über beide Belege und beide Labelzuweisungen. B_D bezeichnet die Differenz der Aberkennungswahrscheinlichkeiten für einen früher gegenüber einem später etikettierten Beleg, über die inhaltlich festgehaltenen Belege, beide randomisierten Labelzuweisungen und tatsächlichen Darbietungsfolgen gemittelt. Für jeden Kontrollvergleich C aus U und N werden sowohl der zusätzliche Fehler R_T minus R_C als auch dessen gerichtete Labelbindung B_T minus B_C geschätzt. Die zweite Größe ist die Wechselwirkung von Darstellungsform und Labelzuordnung; ein in allen Darstellungen gleicher Labeleffekt erfüllt diese Vorhersage nicht. Entscheidungsschwellen: Für alle vier notwendigen Kontraste werden vor der

Hauptstudie positive praktisch relevante Mindestwirkungen sachlich begründet und festgelegt, insbesondere eigenständige Grenzen für T gegenüber N. Es werden hier weder Zahlen eingesetzt noch vorhandene empirische Begründungen behauptet. Stichprobengröße, Modell, Fehlerkontrolle und Intervallverfahren werden auf diese Konjunktion und auf die Abhängigkeiten wiederholter Urteile innerhalb von Personen und Aufgaben ausgelegt. Unterstützung der präzisierten Vorhersage verlangt, dass die nach diesem Verfahren bestimmten unteren Intervallgrenzen die jeweiligen Mindestwirkungen erreichen oder überschreiten; liegt eine obere Intervallgrenze darunter, scheitert mindestens ein notwendiger Bestandteil. Alle übrigen Intervallkonstellationen bleiben für die Konjunktion unentschieden. Sehr kleine positive Effekte können damit die praktische Mindestvorhersage verfehlen, ohne dass jeglicher Effekt ausgeschlossen wäre. Durchführung und Reichweite: Kriterien ungültiger Durchführung werden vorab auf Studienebene festgelegt; alle zugewiesenen Personen bleiben in der primären Analyse, insbesondere unabhängig von nachgelagerten Verständnisantworten. Der Umgang mit fehlenden Antworten wird vorab bestimmt. Absolute Kontraste, Unsicherheit und Sensitivität gegenüber tatsächlicher Expositionsfolge werden berichtet. Nachgelagerte Erläuterungen dienen ausschließlich ergänzender Beschreibung, weder der nachträglichen Fallauswahl noch dem Mechanismusnachweis. Ein von der Grafik veränderter Lese- oder Aufmerksamkeitsverlauf kann zum Verfahrenseffekt gehören. Eine Variation zugeschriebener Autorität und ein unterscheidender Vergleich innerer Modelle gehören in gesonderte Anschlussstudien; der Hauptversuch behauptet keine unabhängig davon identifizierte Wirkung reiner Zeitlichkeit und gilt nicht für echte Rücknahmen oder veränderliche Sachverhalte.

Präzisierung der Zielgrößen. R_D ist die Rate ausdrücklicher Aberkennungen über beide Belege und beide Labelzuweisungen; B_D ist die Differenz der Aberkennungswahrscheinlichkeiten früher versus später etikettierter Belege. Primär: R_T−R_U, R_T−R_N, B_T−B_U, B_T−B_N. Die B-Kontraste beziehen sich auf denselben vorab festgelegten Aufgabenmix und dieselben Zuweisungsgewichte wie die R-Kontraste.

Was wäre der widerlegende Befund? Im vorab festgelegten Aufgabenbereich scheitert die auf praktisch relevante Mindestwirkungen präzisierte Vorhersage bei gültiger Durchführung, sobald die obere Grenze eines nach vorab bestimmter Fehlerkontrolle berechneten Unsicherheitsintervalls für mindestens einen der beiden Darstellungsvergleiche oder dessen gerichtete Bindung an den Labeltausch unter der jeweils begründeten Mindestwirkung liegt. Breite Intervalle und fehlende statistische Signifikanz entscheiden den Test nicht; solange Materialvalidierung, Mindestwirkungen und Präzisionsplanung fehlen, ist seine Durchführung noch nicht vorbereitet.

Ein gut präzisiertes Ergebnis kann kleine Effekte zulassen und trotzdem die vorher festgelegte praktische Mindestvorhersage widerlegen. Die Schwellen dürfen weder aus den späteren Hauptdaten gewählt noch nach einem gescheiterten Direktvergleich abgesenkt werden. Unterhalb der Vorbereitungsanforderungen wird kein entscheidendes Experiment behauptet.

Neuheits- und Übertragungsgrenze. In der begrenzten heutigen Recherche wurde kein direkter Test genau dieser menschlichen Vier-Kontraste-Vorhersage gefunden. Ein Labeltausch bei gleichen Texten wurde bereits von Oh et al. 2026 für zeitliche Fragen an Sprachmodelle untersucht; die Versuchsidee ist daher nicht insgesamt neu.

Offene empirische Fragen

• #verzweigung-offen-empirie-geltungsaberkennung-material — Material und Antwortformat unabhängig pilotieren: fortbestehende Begründungsverwendbarkeit, bloße Auswahl und zugeschriebene Rücknahme müssen ohne suggestiven Unterricht unterscheidbar sein.

• #verzweigung-offen-empirie-chronologie-direktkontraste — Für beide direkten Darstellungsvergleiche und beide gerichteten Labelbindungen praktisch begründete Mindestwirkungen, Ausfallregeln und

gemeinsame Präzisionsplanung entwickeln.

• #verzweigung-offen-empirie-mensch-modell-erfassungslabels — Die bei Oh et al. 2026 untersuchte Metadatenintervention begrifflich und methodisch von menschlicher Geltungsaberkennung abgrenzen; ein späterer Vergleich wäre eine eigene Studie, kein Ersatz für H1.

Empirie-Score

Score: 6/10

Begründung: Die drei Konsequenzen lassen sich in vier konkrete Kontraste übersetzen, und verwandte Literatur begrenzt plausible Alternativerklärungen. Der spezielle Gegenstand ist jedoch noch nicht materialvalidiert; ohne festgelegte Mindestwirkungen, gemeinsame Präzisionsplanung und eigene Daten liegt die Anschlussfähigkeit unter einem bereits teilweise empirisch erschlossenen Prüfpfad. Der Score ist unabhängig von der Neun-Kriterien-Bewertung und kein Bestätigungsgrad.

Rechercheumfang: Gezielte Websuche am 02.11.2026 nach Zeitstempeln, Glaubwürdigkeit, Aktualität, chronologischer Darstellung und Beleggültigkeit; keine systematische Übersichtsarbeit. Originalquellen und tatsächliche Zugriffstiefe stehen im Quellenstudium. Die neue Metadatenstudie, die Feldanalyse und Min wurden erst nach der internen Synthese gefunden.

Externe Prüfung — Literatur und getrennte Gegenprüfungen

Externe Prüfung bedeutet hier aktuelle wissenschaftliche Quellen und getrennte Prüfaufträge. Alle Textprüfungen nutzten GPT-6 Astra/max; es liegt weder eine menschliche Begutachtung noch ein unabhängiger Modellanbieter vor. Die Falsifikation erhielt nur die finale Hypothese und den Schwellentest, Hacking nur die finale Hypothese.

Stage 1 — Originalitätsprüfung (GPT-6 Astra/max, Websuche)

Anschlussfähigkeit (was ist bekannt)

Die finale Fassung liegt an der Schnittstelle von Reihenfolgeeffekten, Informationsvisualisierung, Belegbewertung und der Interpretation zeitlicher Metadaten. Die allgemeine Einsicht, dass eine Anordnung Bedeutungen verändert, ist durch die Diagrammexperimente von Heiser und Tversky (2006) vorweggenommen. Patrick et al. (2012) und Xu und Kwok (2019) liefern Anschlüsse zur Darbietungsfolge; Rottman und Zhang (2025) zeigen, warum zeitliche Anpassung an tatsächlich veränderte Zusammenhänge von einem Fehler unterschieden werden muss. Die Widerrufsliteratur bei Ecker et al. (2015) und Hey et al. (2026) behandelt andere Zielgrößen und tatsächliche Rücknahmen. Insbesondere Eckers Bedingungen ohne Rücknahme rechtfertigen keine allgemeine Behauptung automatischer Ersetzung.

Zwei neu gefundene Quellen liegen näher. Norris et al. (2019) analysieren bereits, wie Zeitstempel, Glaubwürdigkeit und sachlicher Kontext in gemeinsamer Informationsarbeit konkurrieren. Oh et al. (2026), ein am 23. Oktober eingereichter Preprint, verwenden bereits einen Tausch von Erfassungsdaten bei festgehaltenen Revisionstexten und Revisionszeiten. Damit wird gerade die methodische Grundidee kontrolliert veränderter Erfassungsmetadaten bei gleichen Texten vorweggenommen. Die Ergebnisse betreffen Sprachmodelle und zeitabhängig unterschiedliche richtige Antworten; sie sind kein Nachweis menschlicher Geltungsaberkennung. Min (2023) ergänzt eine Grenze pauschaler Aktualitätsheuristiken: Ihre Wirkung hängt davon ab, welche Bedeutung Zeit im jeweiligen Bewertungsfeld besitzt.

Originalitätskern (was ist neu)

Als möglicher eigener Beitrag verbleibt die Kombination aus Menschen als Untersuchungspersonen, zwei sachlich unabhängigen und fortgesetzt gültigen Belegen, zeitlicher Grafik gegenüber zwei Kontrollen und ausdrücklicher Aberkennung weiterer Begründungsverwendbarkeit als Zielgröße. Die Hypothese behauptet zusätzlich, dass gerade der Mehrfehler gegenüber jeder Kontrolle dem früheren Label folgt. In den heute gelesenen Originalquellen und der gezielten Nachsuche wurde kein direkter Test dieser gesamten Konstellation gefunden. Diese Suchgrenze ist kein Prioritätsnachweis und keine Behauptung, dass einzelne Bestandteile neu seien.

Die sachliche Differenz zu Oh et al. muss im späteren Studienprotokoll präzise bleiben: Es geht weder um die Auswahl einer historisch richtigen Revision noch um ein allgemeines Sprachmodellproblem. Auch der Schritt von der ursprünglichen Widerrufsbehauptung zur allgemeineren Geltungsaberkennung erweitert die Zielgröße und reduziert den theoretischen Anspruch. Eine neu benannte Kombination rechtfertigt daher keine hohe Originalitätswertung ohne nachgewiesenen zusätzlichen Erkenntnisgewinn. Empfehlung für die externe Neubewertung: Originalität gegenüber dem internen Wert 6 absenken; den spezifischen menschlichen Kontrollvergleich als begrenzten, ungetesteten Forschungsbeitrag erhalten. Die empirische Vorhersage wird durch verwandte Literatur weder bestätigt noch erledigt.

Recherche am 02.11.2026: aufbauend auf der Empirie-Brücke zusätzliche Suchkombinationen zu timeline/evidence/superseded/human judgment, timestamps/invalid/evidence/psychology sowie temporal/diagram/validity/judgment. Die Treffer wurden auf sachliche Nähe geprüft; unpassende medizinische, juristische und technische Anwendungen werden nicht als Belege verwendet. Kein systematisches Review, keine vollständige Zitationskettenanalyse. Die beiden stärksten neuen Vorarbeiten wurden erst nach dem internen Panel gefunden und nicht rückwirkend seinen Stimmen zugeschrieben.

Quellenliste

- John Patrick; Lewis Bott; Phillip Morgan; Sophia L. King (2012). Out of sequence communications can affect causal judgement. Thinking & Reasoning. DOI: 10.1080/13546783.2012.658240.

- Xinming Xu; Sze Chai Kwok (2019). Temporal-order iconicity bias in narrative event understanding and memory. Memory. DOI: 10.1080/09658211.2019.1622734.

- Benjamin M. Rottman; Yiwen Zhang (2025). Learning about causal relations that change over time: primacy and recency over long timeframes in causal judgments and memory. Cognitive Research: Principles and Implications. DOI: 10.1186/s41235-025-00614-9.

- Ullrich K. H. Ecker; Stephan Lewandowsky; Candy S. C. Cheung; Murray T. Maybery (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. Journal of Memory and Language 85, 101–115. DOI: 10.1016/j.jml.2015.09.002.

- Carolin V. Hey; Marie Luisa Schaper; Ute J. Bayen (2026). The continued influence effect and situation models: Evidence for the model-updating account. Memory & Cognition, veröffentlicht 16.09.2026. DOI: 10.3758/s13421-026-01918-x.

- Julie Heiser; Barbara Tversky (2006). Arrows in Comprehending and Producing Mechanical Diagrams. Cognitive Science 30(3), 581–592. DOI: 10.1207/s15516709cog0000_70.

- Minsu Oh; Seunghui Jwa; Soo Kyun Kim; Yeo-Chan Yoon (2026). Collection-Date Assignment Redirects Answers in Temporal QA. arXiv:2610.32191v1, eingereicht 23.10.2026; Preprint. DOI: 10.48550/arXiv.2610.32191.

• Wendy Norris; Amy Voida; Leysia Palen; Stephen Voida (2019). Is the Time Right Now? Reconciling Sociotemporal Disorder in Distributed Team Work. Proceedings of the ACM on Human-Computer Interaction 3, CSCW, Article 98, 29 Seiten. DOI: 10.1145/3359200.

• Bora Min (2023). Because it is brand new! Recency heuristic for product innovativeness evaluation. International Journal of Consumer Studies 47(3), 1023–1041; online seit 12.11.2022. DOI: 10.1111/ijcs.12885.

Stage 2 — Falsifikationsversuch (frischer GPT-6-Astra/max-Kontext)

Falsifikations-Audit

Die Behauptung ist prinzipiell falsifizierbar, gegenwärtig jedoch nicht ausführungsreif. Sie beansprucht vier notwendige Mindestwirkungen: zusätzliche ausdrückliche Aberkennungen gegenüber U und N sowie jeweils eine zusätzliche Bindung an das frühere Erfassungslabel. Damit kann bereits das präzise Scheitern eines Bestandteils die gesamte Konjunktion widerlegen. Ein kleiner positiver Effekt widerlegt dabei weder jede denkbare Wirkung noch rettet er die beanspruchte praktische Relevanz.

Der stärkste offene Einwand betrifft die noch unbegründeten Schwellen. Solange ihre Größen fehlen, ist nicht festgelegt, welches Ergebnis als hinreichende Wirkung gelten soll. Beliebig kleine nachträgliche Grenzen würden die These gegen schwache Befunde schützen; unrealistisch große Grenzen könnten ihr Scheitern künstlich herstellen. Für jeden Kontrast muss deshalb unabhängig von den Hauptdaten erklärt werden, welche zusätzliche Fehlerrate beziehungsweise Richtungsverschiebung praktisch etwas bedeutet. Die vier Grenzen müssen gemeinsam mit möglichen Antwortwahrscheinlichkeiten vereinbar sein.

Die Materialvalidierung birgt eine zweite Rettungsmöglichkeit: Aufgaben dürfen nicht deshalb als ungeeignet ausscheiden, weil sie keinen erwarteten Chronologieeffekt zeigen. Vorab zu prüfen sind Verständlichkeit, fortbestehende Begründungsverwendbarkeit, Informationswert und die Unterscheidbarkeit der Antwortkategorien. Zu fixieren sind ferner Aufgabenbereich, Auswahlverfahren und Gewichtung. Nach der Hauptstudie dürfen weder ein ungünstiges Aufgabenmuster noch ein schwacher T–N-Kontrast zur nachträglichen Verengung des ursprünglichen Geltungsbereichs führen.

B misst eine gerichtete Differenz im Mittel. Es beweist keinen vollständigen Fehlerwechsel bei jedem einzelnen Belegpaar. Bei identischer Gewichtung gilt für den zusätzlichen Fehler am früher etikettierten Beleg $\Delta R + \Delta B/2$ und am später etikettierten Beleg $\Delta R - \Delta B/2$. Positive R- und B-Kontraste erlauben daher zusätzliche Fehler an beiden Belegen. Diese Reichweite sollte ausdrücklich festgehalten werden; die vorhandene Vorhersage darf weder individualisiert noch auf ausschließliche Schädigung des früheren Belegs erweitert werden.

Konkrete Falsifikationskandidaten

Die folgenden Befundmuster sind vorgeschlagene Gegenproben, keine berichteten Versuchsergebnisse. Geeignete Daten liefert die bereits entworfene randomisierte Hauptstudie.

Erstens: T erzeugt gegenüber U zusätzliche Aberkennungen, gegenüber N jedoch keinen hinreichenden Zusatz. Liegt die obere Intervallgrenze von $R_T - R_N$ unter δ_{TN}, scheitert die These trotz eines eindrucksvollen T–U-Unterschieds. Der zeitneutralen Verbindung darf anschließend nicht allein wegen dieses Befunds ihre Eignung abgesprochen werden; deren vorab festgelegte Validierungskriterien bleiben maßgeblich.

Zweitens: T erhöht die Fehler insgesamt, doch der zusätzliche Fehler folgt dem Labeltausch nicht hinreichend. Ein präzise zu kleiner oder gegenläufiger B-Kontrast genügt. Derselbe Früher-Label-Effekt in allen drei Darstellungen bestätigt die geforderte Wechselwirkung gerade nicht.

Drittens: Der Labeltausch verteilt Aberkennungen zwischen den Belegen um, ohne die gesamte Fehlerrate ausreichend anzuheben. Dann kann die gerichtete Labelbindung vorhanden sein, während ein notwendiger R-Kontrast scheitert. Eine nachträgliche Umdeutung zur bloßen Fehlerumverteilung wäre eine andere Behauptung.

Viertens: Unterschiede hängen vorwiegend an der tatsächlichen Erstexposition oder bestimmten Beleginhalten. Das widerlegt die These erst, wenn nach der geplanten Ausbalancierung mindestens ein notwendiger Kontrast präzise unter seiner Schwelle liegt. Ein zusätzlicher Reihenfolgeeffekt allein genügt nicht. Ebenso sind missverständliche Antwortformulierungen zunächst ein Messproblem, kein empirischer Gegenbeweis.

Schwellentest

Der entscheidende Test bleibt ein einziger vorab festgelegter Versuch mit T, U und N, unabhängig randomisierten Erfassungslabels und tatsächlichen Erstexpositionsfolgen. Nach abgeschlossener, separat durchgeführter Validierung werden Aufgabenmix, Antwortkodierung, Gewichte und alle vier positiven Schwellen festgeschrieben. Ein Pilot darf die Messbarkeit prüfen; die Hauptvorhersage darf nicht durch Auswahl der effektstärksten Aufgaben vorbereitet werden.

Ausgewertet werden alle zugewiesenen Personen nach dem vorab bestimmten Verfahren. Fehlende Antworten, Studienabbruch und technische Ungültigkeit erhalten überprüfbare Regeln. Nachgelagerte Verständnisantworten erlauben keine erfolgsabhängige Fallauswahl. Personen- und Aufgabenabhängigkeiten müssen im Unsicherheitsmodell berücksichtigt werden.

Die Schärfung besteht in simultan abgesicherten Intervallen für genau die vier primären Kontraste und einer Präzisionsplanung für die gemeinsame Entscheidung. Simulationsszenarien müssen insbesondere berücksichtigen, dass nur ein Bestandteil die Mindestwirkung verfehlt. Vier getrennt günstig dimensionierte Einzeltests reichen dafür als Planung nicht aus.

Erreichen alle unteren Grenzen ihre Schwellen, bewährt sich die präzisierte Vorhersage im festgelegten Bereich. Unterschreitet wenigstens eine obere Grenze ihre Schwelle, scheitert sie dort. Alles Übrige bleibt offen. Unentschiedenheit darf weder als Bestätigung noch als Anlass zur rückwirkenden Schwellenänderung verbucht werden. Ein bestandener Test weist weiterhin keinen bestimmten inneren Mechanismus oder zugeschriebenen Rücknahmeakt nach.

Stage 3 — Schul-fremde Begutachtung: Hacking (frischer GPT-6-Astra/max-Kontext)

Aus einer KI-rekonstruierten, an Hacking orientierten Perspektive überzeugt zunächst, dass die Hypothese eine bestimmte Anordnung von Zeichen mit beobachtbaren Antworten verbindet und ihre Reichweite begrenzt. Die Trennung zwischen Belegauswahl, weiterer Begründungsverwendbarkeit und zugeschriebener Rücknahme verhindert mehrere begriffliche Kurzschlüsse. Eine spezifisch deutsche Voraussetzung lässt sich aus diesem Entwurf nicht belegen. Sichtbar wird jedoch eine theoretische Entscheidung: Die experimentelle Aufgabenregel soll festlegen, welche Verwendung eines Belegs berechtigt ist, und daran wird eine Antwort als Fehler bemessen. Damit entsteht ein untersuchbarer Gegenstand innerhalb einer eingerichteten Praxis. Ob derselbe Gegenstand außerhalb dieser Praxis als unbegründete Geltungsaberkennung wiederkehrt, bleibt eine zusätzliche Frage. Gerade die Wörter „gültig“, „Begründung“ und „dürfen“ verlangen Aufmerksamkeit: Ein Dokument kann eine Behauptung stützen, für einen Arbeitsvorgang entbehrlich sein oder nach einer institutionellen Vorschrift unzulässig werden. Der Entwurf unterscheidet solche Möglichkeiten ausdrücklich; seine Tragfähigkeit hängt deshalb daran, dass diese Unterscheidungen auch die tatsächlichen Antwortverwendungen der Versuchspersonen treffen. Mein schärfster Einwand betrifft die Vergleichbarkeit des gemessenen Fehlers zwischen den Darstellungen. Eine zeitlich geordnete Verbindung kann eine Konvention der Ersetzung nahelegen und dadurch verändern, wie jemand die Frage nach weiterer Verwendbarkeit versteht. Dann wäre eine identische Antwortformulierung unter verschiedenen Bedingungen möglicherweise Ausdruck verschiedener Urteile. Randomisierung und Labeltausch identifizieren einen Effekt der Darstellung auf Antworten, entscheiden diese begriffliche Frage aber noch nicht. Dass der Entwurf keinen Mechanismus reiner Zeitlichkeit beansprucht, ist deshalb richtig; das Problem reicht jedoch bis zur Benennung seines primären Endpunkts. Die konkrete Abhilfe liegt in einer unabhängigen Vorprüfung unter sämtlichen Darstellungsbedingungen: An zusätzlichem Material sollten neutral erhobene Paraphrasen und getrennte Fälle von fortbestehender Begründung, bloßer Entbehrlichkeit und ausdrücklicher Unzulässigkeit zeigen, ob die Antwortkategorien vergleichbar gebraucht werden. Die Auswertung benötigt vorab festgelegte Kriterien und gegenüber den Bedingungen möglichst verblindete Kodierung. Solche Befunde dürfen keine nachträgliche Auswahl passender Personen begründen. Gelingt diese Prüfung, erhält der Hauptversuch eine präzisere Aussage darüber, welcher regelbezogene Fehler zunimmt; misslingt sie, bleibt zunächst ein Darstellungseffekt auf sprachliche Antworten. Als ausdrücklich gesonderten Folgeansatz schlage ich eine historische Untersuchung von Ablagepraktiken vor: Wann und durch welche Arbeitsregeln wird ein Erfassungsdatum zum Zeichen verminderter Verwendbarkeit? In vergleichenden Studien könnten tatsächliche institutionelle Regeln und bloße grafische Konventionen getrennt variiert werden. Dabei ließe sich untersuchen, ob die Einführung einer Klassifikation wie „überholter Beleg“ das weitere Handeln mit Dokumenten verändert und dadurch jene Nutzungsmuster hervorbringt, die die Klassifikation anschließend zu beschreiben scheint. Ein solcher möglicher Rückkopplungsprozess wäre eigenständig nachzuweisen; er ersetzt weder die vier Kontraste des Haupttests noch dessen noch ausstehende Materialvalidierung.

Konsequenzen für die Endfassung

Die externe Bewertung beträgt 61/90 (Rohsumme 67.8; Faktor 0,9), gegenüber intern 64 und initial 68. Originalität sinkt 6→5 wegen naher publizierter Metadaten- und Zeitstempelarbeiten; Falsifizierbarkeit 8→7 wegen noch nicht verbindlich begründeter, gemeinsam zu planender Mindestwirkungen; Publikationsmöglichkeit 7→6 wegen des engeren zusätzlichen Beitrags und fehlender Daten. Begriffliche Klarheit bleibt 8: Die Definitionen sind getrennt, aber ihre vergleichbare Verwendung muss unter sämtlichen Darstellungen vorab validiert werden. Ein präziserer Entwurf ist damit kein bestätigter oder fertiger Versuch. Die Einwände ändern weder rückwirkend die internen Gutachten noch den gewählten Hauptlinienprozess; die Scores der nur kurz geprüften Nebenlinie sind mit dieser vertieften externen Bewertung nicht gleichzusetzen.

Die Labelbindung ist eine gerichtete mittlere Verschiebung, kein behaupteter individueller Fehlerwechsel und keine ausschließliche Schädigung des früheren Belegs. Bei identischer Gewichtung gelten für zusätzliche Fehler am früheren und späteren Beleg $\Delta R+\Delta B/2$ beziehungsweise $\Delta R-\Delta B/2$; beide können positiv sein.

Die unabhängige Materialvorprüfung muss alle drei Darstellungen erfassen. Neutrale Paraphrasen und eigene Fälle für Begründungsverwendbarkeit, Entbehrlichkeit und Unzulässigkeit werden nach vorab fixierten Kriterien möglichst bedingungsverblindet kodiert.

Aufgaben und Personen werden nicht nach dem Auftreten des gewünschten Effekts ausgewählt. Schwellen, Aufgabenmix, Gewichte, Ausfälle und Studienungültigkeit werden vor Hauptdaten festgelegt; die vier Mindestwirkungen müssen gemeinsam möglich sein.

Die gemeinsame Präzisionsplanung berücksichtigt insbesondere Szenarien, in denen nur einer der notwendigen Kontraste scheitert. Separate günstige Einzeltestplanungen genügen dafür nicht.

Scheitert die Vergleichbarkeit der Antwortkategorien, liegt gegebenenfalls ein Effekt auf sprachliche Antworten vor, aber kein gültiger Test der als Geltungsaberkennung bezeichneten Zielgröße. Das Material ist dann vor einer Hauptstudie neu zu entwickeln; ein bereits beobachtetes ungünstiges Hauptresultat darf nicht umetikettiert werden.

Historische und institutionelle Vergleiche von Erfassungsdaten als Statuszeichen sind eine gesonderte Folgestudie; sie ersetzen keinen gescheiterten Hauptvergleich.

Quellenstudium der Hauptlinie

Gezielte Prüfung wissenschaftlicher Originalquellen mit dokumentierter Teillektüre, keine systematische Übersichtsarbeit und keine Datenreanalyse. Verwandte Befunde begrenzen oder motivieren den Prüfentwurf; ein direkter empirischer Nachweis der heutigen Hypothese liegt nicht vor. Volltexte werden verlinkt und nicht in den Vault kopiert.

John Patrick; Lewis Bott; Phillip Morgan; Sophia L. King (2012)

John Patrick; Lewis Bott; Phillip Morgan; Sophia L. King (2012). Out of sequence communications can affect causal judgement. Thinking & Reasoning. DOI: 10.1080/13546783.2012.658240.

Gelesen: Originalabstract und Metadaten im Universitätsrepositorium heute geöffnet; dort kein Volltext.

Anschluss: Drei Experimente zeigen geänderte Ursachenzuschreibung, wenn Ereignisse einer Ursache kausal geordnet und jene einer anderen verschränkt präsentiert werden.

Grenze: Keine reine Arbeitsprovenienz-Manipulation, keine allgemeine Aussage über Abschlussrichtigkeit.

Vault: Patrick et al. 2012 - Darbietungsfolge und Kausalurteil.

Xinming Xu; Sze Chai Kwok (2019)

Xinming Xu; Sze Chai Kwok (2019). Temporal-order iconicity bias in narrative event understanding and memory. Memory. DOI: 10.1080/09658211.2019.1622734.

Gelesen: Originalabstract und Metadaten im Duke-Autorenrepositorium, Websuchindex heute; kein Volltext.

Anschluss: Erzählte Videoreihenfolge beeinflusst erschlossene Chronologie. Explizite Chronologieinformationen verändern die Verzerrungsrichtung.

Grenze: Erinnerung und Reihenfolgeurteil sind keine Prüfung richtiger Schlussfolgerungen.

Vault: Xu und Kwok 2019 - Erzählreihenfolge und Zeiturteil.

Benjamin M. Rottman; Yiwen Zhang (2025)

Benjamin M. Rottman; Yiwen Zhang (2025). Learning about causal relations that change over time: primacy and recency over long timeframes in causal judgments and memory. Cognitive Research: Principles and Implications. DOI: 10.1186/s41235-025-00614-9.

Gelesen: Originalabstract, Einleitungs-, Methoden- und Schlussauszüge der Verlagsseite im Websuchindex heute; PubMed39982529 geprüft. Kein vollständiger Ergebnisteil und keine Datenreanalyse.

Anschluss: Kurzzeitaufgabe und 24-Tage-Verfahren untersuchen wechselnde Kausalstärke. Zeitrahmen beeinflusst Aktualitäts- und Erinnerungseffekte; erkannte Kontingenzänderung darf nicht mit Bias gleichgesetzt werden.

Grenze: Das Ergebnis begrenzt pauschale Reihenfolgeeffekte; weder Arbeitsprovenienz noch institutionelle Vertrauensaussetzung wurden getestet.

Vault: Rottman und Zhang 2025 - Kausallernen über wechselnde Zeiträume.

Ullrich K. H. Ecker; Stephan Lewandowsky; Candy S. C. Cheung; Murray T. Maybery (2015)

Ullrich K. H. Ecker; Stephan Lewandowsky; Candy S. C. Cheung; Murray T. Maybery (2015). He did it! She did it! No, she did not! Multiple causal explanations and the continued influence of misinformation. Journal of Memory and Language 85, 101–115. DOI: 10.1016/j.jml.2015.09.002.

Gelesen: Autorenmanuskript: Highlights, Abstract, Einleitung, Methodenteile beider Experimente und Ergebnisdiskussion von Experiment 2 selektiv gelesen; Manuskriptseiten 10–13 und 20–26 vertieft. DOI und Publikationsdaten im UWA-Autorenrepositorium geprüft. Keine Datenreanalyse.

Anschluss: Darbietungsposition und tatsächliche zeitliche Nähe sind zu unterscheiden. Bei kontrollierter Aktualität wirkte die Rücknahme beider Erklärungen ähnlich. Ohne Rücknahme zeigte sich keine signifikante Bevorzugung der zweiten Erklärung; die gepoolte Bayes-Analyse stützte keinen Unterschied nur schwach.

Grenze: Ein pauschaler automatischer Ersetzungseffekt wird damit fraglich. Grafische Verknüpfung sachlich irrelevanter Erfassungslabels ist nicht untersucht; die Befunde bestätigen oder widerlegen H1 deshalb nicht direkt.

Vault: Ecker et al. 2015 - Aktualität und Rücknahme von Erklärungen.

Carolin V. Hey; Marie Luisa Schaper; Ute J. Bayen (2026)

Carolin V. Hey; Marie Luisa Schaper; Ute J. Bayen (2026). The continued influence effect and situation models: Evidence for the model-updating account. Memory & Cognition, veröffentlicht 16.09.2026. DOI: 10.3758/s13421-026-01918-x.

Gelesen: Verlagsvolltext: Abstract, Einleitung, Design, Teilnehmende, Ablauf sowie zentrale Diskussions- und Offenlegungsabschnitte selektiv gelesen. Keine vollständige Tabellenprüfung oder Datenreanalyse.

Anschluss: Nach tatsächlichen Widerrufen untersucht ein faktorielles Experiment Glaubwürdigkeit und den Zeitpunkt ihrer Mitteilung. Erfasste Situationsmodelle vermitteln im statistischen Modell Teile der

Glaubwürdigkeitseffekte; zusätzliche Effekte bleiben erklärungsbedürftig. 478 Personen wurden erfasst, 435 nach Ausschlüssen analysiert.

Grenze: Keine Prüfung bloß fingierter Widerrufe. Nicht präregistriert; statistische Mediation ist für H1 kein identifizierter innerer Mechanismus. Tatsächliche Datenerhebung 2023, Veröffentlichung 2026.

Vault: Hey et al. 2026 - Situationsmodelle nach Widerrufen.

Julie Heiser; Barbara Tversky (2006)

Julie Heiser; Barbara Tversky (2006). Arrows in Comprehending and Producing Mechanical Diagrams. Cognitive Science 30(3), 581–592. DOI: 10.1207/s15516709cog0000_70.

Gelesen: Original-PDF auf der Columbia-Fakultätsseite geöffnet; Abstract, Beschreibung des ersten Experiments, Ergebnisse und Diskussion des zweiten Experiments sowie allgemeine Diskussion selektiv gelesen. Publikationsmetadaten über den Verlagsindex geprüft; keine Datenreanalyse.

Anschluss: In zwei Experimenten wurden mechanische Diagramme mit Pfeilen eher funktional und ohne Pfeile eher strukturell beschrieben; bei funktionalen Textbeschreibungen zeichneten Teilnehmende häufiger Pfeile. Grafische Verbindungen können damit dynamische und kausale Beziehungen vermitteln.

Grenze: Die allgemeine bedeutungsverändernde Leistung von Pfeilen ist eine bekannte Vorarbeit, kein neuer Beitrag von H1. Die Studie prüft weder sachlich irrelevante Erfassungslabels noch falschen Geltungsentzug weiterhin gültiger Belege; Pfeilwirkung und spezifischer Chronologieeffekt sind getrennt zu testen.

Vault: Heiser und Tversky 2006 - Pfeile und funktionales Diagrammverständnis.

Minsu Oh; Seunghui Jwa; Soo Kyun Kim; Yeo-Chan Yoon (2026)

Minsu Oh; Seunghui Jwa; Soo Kyun Kim; Yeo-Chan Yoon (2026). Collection-Date Assignment Redirects Answers in Temporal QA. arXiv:2610.32191v1, eingereicht 23.10.2026; Preprint. DOI: 10.48550/arXiv.2610.32191.

Gelesen: Originalabstract, Einleitung, Aufgaben- und Interventionsbeschreibung sowie Ergebnis- und Reichweitenabschnitte 4–6 im arXiv-HTML selektiv gelesen; keine vollständige Appendix-Prüfung, keine Datenreanalyse, keine unabhängige Replikation.

Anschluss: Bei festgehaltenen Revisionstexten und Revisionszeiten kann der Tausch separater Erfassungsdaten die Antwortauswahl mehrerer Sprachmodelle stark verändern. Effekte hängen unter anderem von Modell und Nähe zum Fragedatum ab; API-Modelle zeigen in den geprüften Konfigurationen deutlich kleinere Effekte.

Grenze: Die Labeltausch-Idee bei unverändertem Belegmaterial ist damit bereits publiziert. Untersucht werden Sprachmodelle, unterschiedliche historische Sachstände und Antwortauswahl, nicht Menschen mit zwei weiterhin gültigen unabhängigen Belegen, grafischen Zeitverbindungen und ausdrücklicher Geltungsaberkennung. Preprint, keine bestätigte Übertragung auf H1.

Vault: Oh et al. 2026 - Erfassungsdatum und Antwortauswahl.

Wendy Norris; Amy Voida; Leysia Palen; Stephen Voida (2019)

Wendy Norris; Amy Voida; Leysia Palen; Stephen Voida (2019). Is the Time Right Now? Reconciling Sociotemporal Disorder in Distributed Team Work. Proceedings of the ACM on Human-Computer Interaction 3, CSCW, Article 98, 29 Seiten. DOI: 10.1145/3359200.

Gelesen: Original-PDF auf der Autorenwebsite: Abstract, Einleitung und Abschnitte zur Zeitordnung, besonders S. 15–18 des PDFs, selektiv gelesen; keine erneute Auswertung der Chatdaten.

Anschluss: Die qualitative Analyse digitaler humanitärer Zusammenarbeit zeigt Konflikte zwischen Zeitstempeln, Glaubwürdigkeit und inhaltlichem Kontext. Ein zunächst widersprüchlich erscheinendes Berichtspaar erweist sich bei Nachprüfung als in beiden Fällen korrekt; bloße Zeitstempel führen hier zur irreführenden Einordnung.

Grenze: Der allgemeine Konflikt zwischen Chronologie und Beurteilung ist bereits dokumentiert. Die Feldstudie isoliert keinen kausalen Grafikeffekt und enthält keine randomisierte Gültigkeitsmessung bei unverändertem Sachverhalt.

Vault: Norris et al. 2019 - Zeitstempel und gemeinsame Informationsprüfung.

Bora Min (2023)

Bora Min (2023). Because it is brand new! Recency heuristic for product innovativeness evaluation. International Journal of Consumer Studies 47(3), 1023–1041; online seit 12.11.2022. DOI: 10.1111/ijcs.12885.

Gelesen: Verlagsvolltext: Abstract, Einleitung, Studie 3 mit Methode und Ergebnisdiskussion sowie allgemeine Diskussion selektiv gelesen; ergänzende Materialien und Daten nicht geprüft.

Anschluss: Aktualitätsangaben verändern unter bestimmten Bedingungen Urteile über Produktinnovativität. In Studie 3 tritt bei Selbsthilfebüchern kein signifikanter Aktualitätseffekt auf, während Aktualität bei Reiseführern sachlich relevant ist.

Grenze: Das ist kein Äquivalenzbeweis und kein Test von Beleggültigkeit. Die Befunde sprechen gegen eine kontextfreie Ableitung der heutigen Hypothese allein aus allgemeiner Aktualitätspräferenz.

Vault: Min 2023 - Aktualitätsheuristik und ihre Grenzen.

Rückkopplung für den nächsten Lauf

Welche zusätzliche Intervention liefert unterschiedliche Vorhersagen für eine veränderte Geltungsannahme und die Anpassung an eine vermutete Antworterwartung, wenn beide dieselbe ausdrückliche Geltungsaberkennung erzeugen können? Die weiterführenden empirischen Fragen sind als eigene Reservoir-Notizen abgelegt. Die Literatur liefert Anschluss und Gegenbefunde; neue Daten, Materialvalidierung und vorab begründete Entscheidungsgrenzen stehen noch aus.

Reservoir und Rückkopplung

- 06 Hypothesentag/Reservoir/Reservoir - Vorleistung eröffnet einen gemeinsamen Zweck 2026-11-02
- 06 Hypothesentag/Reservoir/Reservoir - Wer die Form zerbrechen darf 2026-11-02
- 06 Hypothesentag/Reservoir/Reservoir - Offene Frage empirie-geltungsaberkennung-material 2026-11-02
- 06 Hypothesentag/Reservoir/Reservoir - Offene Frage empirie-chronologie-direktkontraste 2026-11-02
- 06 Hypothesentag/Reservoir/Reservoir - Offene Frage empirie-mensch-modell-erfassungslabels 2026-11-02
- 06 Hypothesentag/Reservoir/Reservoir - Offene Frage institutionelle-datierung-belegstatus 2026-11-02

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.