Theoretical article with a preregistrable research agenda

Why Conspiracy Explanations Can Feel More Complete Than Official Accounts: A Concealment-Weighting Model

Shaobang Sun, Ph.D.
CupTalks.com

Abstract

Why can a hidden-cause account feel more complete than an institutional explanation despite weak evidence? We propose a conditional account: an alleged cover-up may lead recipients to discount a failed search for records more than the specified concealment mechanism warrants. A failed search should weaken a hidden account in proportion to the chance that its predicted trace would have survived and been found; when concealment completely prevents that trace, the search is uninformative. The proposed excess discounting is distinct from the legitimate adjustment for concealment. Evidence on a general complexity advantage and on unusually high need for cognition among believers is mixed or contrary, so those claims are not premises of the model. We map the relevant evidence, specify residual-based tests against stipulated, participant-elicited, and neutral-item benchmarks, and compare the account with motivated reasoning and source distrust. Four studies examine paired explanations, evidence updating, discrimination of verified from unsupported plots, and experimental production and reception. The model would be weakened if absence evidence receives its warranted weight.

Keywords: conspiracy theories, explanation, hidden causes, concealment, evidence of absence, source trust, epistemic calibration

Introduction

An official inquiry may conclude that several interacting failures produced a consequential event while leaving some questions unanswered. An alternative account may assign each unanswered question to a concealed group with a clear motive. Many readers find the second account more satisfying. The question is why, and whether the satisfaction tracks anything the evidence supports.

Real cases show why the question cannot be answered by simply trusting or doubting institutions. During the COVID-19 pandemic, claims about concealed origins and suppressed evidence circulated alongside changing public-health explanations. After the 2020 U.S. election, allegations of organized fraud were evaluated amid partisan expectations and conflicting claims. Some covert coordination is also verified by documentary record. These cases differ sharply in evidence and context, and we use them only to illustrate that an account's apparent ability to explain anomalies must be distinguished from its ability to predict independently checkable

observations. Cross-national work cautions against treating distrust as a peculiarity of one event or country: conspiracy belief covaries with country-level corruption, and the social and political setting shapes which alleged plots seem plausible (Alper, 2023; Hornsey et al., 2023).

The present paper concerns perceived credibility, explanatory satisfaction, and the weighting of evidence, not the claim that alternative accounts are generally more accurate or more widely believed. "Official" identifies a source, not a guarantee of truth. Verified conspiracies exist, and each hidden-cause claim deserves assessment on its own evidence (Dentith, 2014; Douglas & Sutton, 2023). The theory therefore concerns how two accounts are compared under uncertainty. It is not a rule to reject allegations of secrecy.

Existing reviews mainly catalog correlates of conspiracy belief: motives, personality, cognitive style, and social context (Biddlestone et al., 2025; Bowes et al., 2023; Douglas et al., 2017). They say little about the comparison process that unfolds when evidence arrives for or against a hidden account. The most direct attempts to explain appeal through explanatory form, namely that more hidden variables or more detail make an account persuasive, have not held up (Marsh & Kleinberg, 2025; Meuer et al., 2021). A narrower mechanism is needed. This paper offers one and states how it could fail. It makes four contributions:

- A normative benchmark for how a failed search should affect belief in a hidden account when concealment is alleged (Figure 1).
- A concealment-weighting account, with observed minus warranted log-odds updating as its primary measure (Figure 2).
- Five hypotheses and an evidence map showing which premises are supported, mixed, contradicted, or untested (Table 1), including a comparison with rival accounts (Table 2).
- A preregistrable research agenda whose designs separate the account from its rivals, with verified and unsupported cases in the stimulus sets.

Felt Completeness and Evidential Constraint

A conspiracy explanation attributes an event to covert coordination among actors. Its defining content invites inferences about inaccessible decisions, motives, and communications (Douglas & Sutton, 2023). Such latent causes can connect disparate observations, which raises what we call *felt completeness*: the perceived share of stated observations that an account explains. They also create a demand for independent evidence. *Evidential constraint* is the degree to which an account's predictions, specified before the test evidence is known, discriminate it from its rival. An account can feel complete and be weakly constrained. We call this dissociation the *hidden-cause gap*. It is a pattern of two separate estimates, high completeness with low constraint, and not a difference score between incomparable scales.

Two routes can open the gap. The first is anomaly-focused comparison. In online discussion of the September 11 attacks, critics of conventional accounts often emphasized apparent anomalies more than they offered a coherent alternative explanation (Wood & Douglas, 2013). Content analysis of paired real-world articles likewise found lower judged falsifiability in conspiratorial texts (Meuer et al., 2023). An audience can then judge an official account against every unresolved detail while judging the alternative by its capacity to suggest that something hidden could explain those details. Both findings come from specific corpora and do not show that the same process governs other events. The second route is the treatment of missing evidence, which we develop below.

The complexity hypothesis, in its unconditional form, does poorly. People often prefer explanations with fewer causes and narrower latent scope (Khemlani et al., 2011; Lombrozo, 2007). Marsh and Kleinberg (2025) found that fact-based alternatives were judged to have broader explanatory and latent scope than conspiracy accounts in their materials, and Meuer et al. (2021) found no credibility gain from merely adding explanatory or threat-related information. We therefore do not claim that more hidden variables make an account more credible. We retain a conditional version (Hypothesis 3): recipients may credit a hidden account for the observations it accommodates while under-charging it for the unsupported auxiliaries that did the accommodating.

A Normative Benchmark for Hidden Accounts

Let H be the hidden-cause account, O its conventional rival, and E the complete observed record. The posterior odds are

$$P(H \mid E) / P(O \mid E) = [P(H) / P(O)] \times [P(E \mid H) / P(E \mid O)].$$

If H leaves actors, tactics, and outcomes adjustable, its likelihood must average over those configurations, weighted by their priors. Fitting the most convenient configuration after inspecting E and treating that fit as the likelihood of H inflates support, because configurations that would have predicted different observations are ignored. Flexibility without concentrated predictive success is not corroboration, although it does not mechanically penalize every hidden account.

Evidence of absence and the role of concealment

A failed search is evidence only relative to a search process and a specified alternative (Hahn & Oaksford, 2007). Let t be the probability that H produces a relevant record, c the probability that such a record is successfully suppressed, and q the probability of detecting a surviving record. Let f be the chance that an incidental lookalike is found under either H or O. We call the latter equality the symmetric-lookalike assumption. Write $\tau = t(1 - c)q$. Assume that

trace creation, suppression, and detection follow these conditional probabilities given H and that an incidental record arises independently of a diagnostic record. Then P(no find | H) = (1 − τ)(1 − f) and P(no find | O) = 1 − f, hence

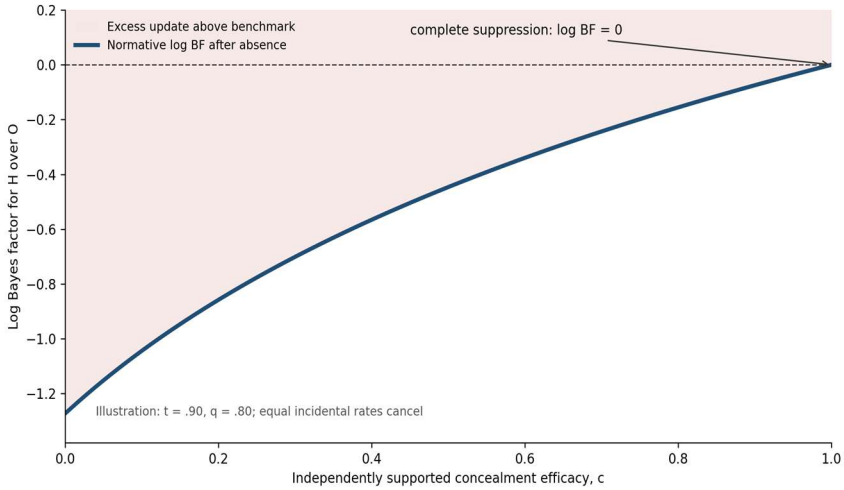
$$BF_{abs} = P(\text{no find} \mid H) / P(\text{no find} \mid O) = 1 - t(1 - c)q.$$

The absence Bayes factor ranges from 1 − tq when c = 0 to 1 when c = 1. Thus a failed search ordinarily weighs against H, and complete suppression makes it uninformative. With t = .90 and q = .80, these endpoints are .28 and 1, respectively; the symmetric-lookalike assumption makes f cancel. For any fixed c below 1, the smallest attainable absence factor is c (at tq = 1). In particular, at c = .5 an absence factor below .5 is impossible, so the illustrative .28 factor at c = 0 cannot be matched in that condition. The matched-item experiment uses factors in the common feasible range; Figure 1 illustrates the general benchmark, not its matched stimulus values. If incidental rates differ under H and O, a separate formula applies (Appendix).

Concealment must be specified before seeing the search outcome. A proposed suppression mechanism has its own prior; inference about H should average over plausible mechanisms rather than select the most favorable c afterward. Documented precedents or observable effects can justify a higher c. The direct contradictions in the proposed experiment are stipulated to be outside the reach of suppression, which makes them comparison items under this model, not universally privileged evidence. A separate appendix clarifies how a secondary trace of concealment would enter the joint likelihood.

Figure 1

Normative Update After a Failed Search Under Alleged Concealment



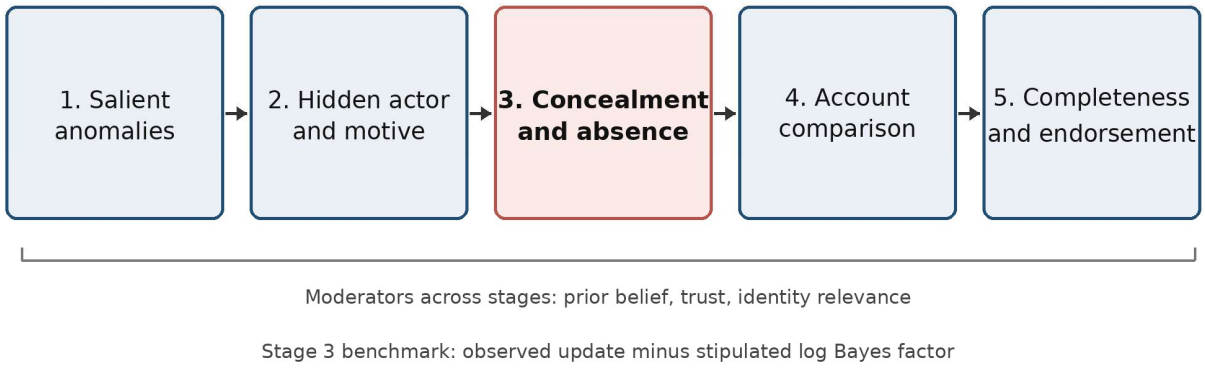
Note. The curve is ln BF_{abs} = ln[1 − t(1 − c)q], with t = .90 and q = .80. It rises from ln(.28) ≈ −1.273 at c = 0 to zero at c = 1. The shaded area marks observed updating beyond the stipulated benchmark, not a finding about participants. This illustrative curve is distinct from the matched experimental stimuli, whose Bayes factors remain in the feasible common range.

The Concealment-Weighting Account

The psychological proposal is selective excess discounting: an allegation of concealment may reduce the evidential weight of a failed search more than the independently warranted reduction allows. A weak negative update is not itself an error, because genuine concealment makes a trace less likely to be found. Under the symmetric-lookalike assumption, complete suppression makes absence neutral. The separate excess-confirmation test compares updating after that failed search with updating after an equally nondiagnostic item that carries no concealment claim. The primary test compares residual belief changes after absent versus direct contradictory evidence. This is not a claim that every conspiracy account is unfalsifiable (Boudry & Hofhuis, 2024).

Figure 2 orders the proposed operations as a heuristic sequence, not strict serial stages. Unresolved details make a hidden actor salient; agency detection, pattern perception, and the expectation that a major event has a major cause make such an actor available (Douglas et al., 2016; Leman & Cinnirella, 2007; van Prooijen et al., 2018). An optional concealment claim changes the expected visibility of evidence. The recipient then compares accounts, and judgments of completeness and belief follow. Prior belief, institutional trust, and identity relevance can influence every stage. The distinctive prediction concerns the third stage.

Figure 2
The Concealment-Weighting Account



Note. Arrows indicate a proposed order of comparison, not established serial causation. The spanning bracket indicates that prior belief, institutional trust, and identity relevance can affect every stage. The benchmark applies at Stage 3.

Parameters

For item i, let ℓ_i be its prespecified log Bayes factor for H over O, u_i the recipient's before-to-after change in log odds, and $r_i = u_i - \ell_i$ the residual update. Under the stated likelihoods, calibrated updating implies $r_i = 0$. The primary analysis compares residuals for absence and

direct contradictions across concealment conditions. A complementary participant-specific benchmark uses the respondent's elicited probabilities of finding the record under H and O; divergence from the stipulated benchmark can then be distinguished from disbelief in the stated frequencies, although post-stimulus elicitation can itself affect responses. A slope-and-intercept model of u_i against ℓ_i across varied items is exploratory because each person has few observations per cell; group-level estimates require crossed participant and scenario effects.

At stipulated complete suppression ($c = 1$), the absence benchmark is $\ell_i = 0$. The excess-confirmation test contrasts mean residual updating after that failed search with mean updating after a neutral item also assigned $\ell_i = 0$, presented with no concealment claim. This controls for a general presentation or response tendency. The within-person contrast and each condition's mean are reported, alongside a personalized likelihood benchmark; a positive raw update by itself does not establish selective confirmation.

Predictions

- Selective down-weighting. Relative to direct contradictory items of matched stipulated diagnosticity, absence items receive less negative weight when concealment is alleged than when it is not, after the normative change in their Bayes factors is taken into account. The condition-by-evidence-type difference in residuals, rather than raw updates, is the primary test.
- Excess confirmation. Under complete suppression and the symmetric-lookalike assumption, the failed search yields a larger positive residual than a neutral, equally nondiagnostic item presented without concealment framing. A positive raw update alone does not satisfy this prediction.
- Trust moderation. The selective residual and excess-confirmation effects are larger under poor institutional competence history or low measured trust; the competence manipulation need not affect every type of evidence.
- Completeness and discrimination. Explicit hidden-agent and concealment content are predicted to raise felt completeness. Any discrimination gain is assessed separately against a preregistered minimum worthwhile change; a null significance test alone cannot establish that there is no gain.

The account is disfavored in this paradigm if absence updates match the stipulated and participant-elicited benchmarks across concealment conditions, if the complete-suppression contrast does not exceed the neutral-item control, or if concealment affects absent and direct evidence comparably. A generic preference for hidden plots can change baseline odds without the predicted evidence-type interaction.

Conditions That Make Hidden Causes Available and Attractive

Several literatures bear on why a hidden account might be generated, attended to, or provisionally accepted. They are best read as moderators and routes to hypothesis generation, not as evidence for the concealment mechanism.

Agency and pattern perception. Hypersensitive agency detection and illusory pattern perception predict conspiracy belief (Douglas et al., 2016; van Prooijen et al., 2018), and a major event can make a major intentional cause feel proportionate (Leman & Cinnirella, 2007). These routes make a hidden actor available but do not show that the actor predicts independent evidence.

Motives and cognitive style. Epistemic, existential, and social motives show modest and heterogeneous associations with endorsement (Biddlestone et al., 2025; Douglas et al., 2017). Need for cognitive closure may heighten attraction to a salient conspiracy claim when a clear official answer is absent (Marchlewska et al., 2018), whereas an earlier small study found no simple trait-level association (Leman & Cinnirella, 2013). Intuitive thinking and fewer draws before deciding in a laboratory task have also been associated with endorsement (Pytlik et al., 2020). Desire for uniqueness has been linked to the appeal of possessing scarce information (Lantian et al., 2017). None of these specifies how absent evidence is processed.

Felt versus demonstrated understanding. An illusion of explanatory depth about politics has been associated with conspiracy endorsement, particularly among political novices and supporters of the losing candidate (Vitriol & Marsh, 2018). This motivates measuring felt completeness separately from demonstrated causal knowledge.

Source trust and vigilance. People monitor communicators for trustworthiness (Sperber et al., 2010), and a well-founded concern about an institution's competence can rationally reduce reliance on its testimony. Yet an institutional error did not automatically raise trust in a competing source in experiments by Dawson et al. (2025). Across countries, higher cognitive reflection has generally correlated with less COVID-19 conspiracy belief (Kantorowicz-Reznichenko et al., 2022). The account therefore predicts a conditional source-by-content effect, not an automatic transfer of credibility. A further caution: endorsement of mutually contradictory conspiracy theories suggests that rejection of the official account, more than coherence of any alternative, can drive acceptance (Wood et al., 2012). This points to source-level distrust as a rival to our mechanism, and Table 2 addresses it.

Hypotheses and Evidence Map

H1 (primary; selective concealment weighting). After an allegation of concealment, recipients give a failed search less negative weight than an equally diagnostic direct contradiction beyond the warranted change in likelihood. Under the symmetric-lookalike assumption, complete

suppression makes absence nondiagnostic; a residual update exceeding that to a neutral, nondiagnostic item is a separate prediction.

H2 (completeness with limited discrimination gain). Explicit hidden-agent content and concealment claims raise felt completeness. We separately estimate any effect on discrimination of verified from unsupported plots, with confidence intervals and a preregistered minimum worthwhile discrimination gain. The dissociation claim is supported only if the completeness effect is positive and the discrimination effect is credibly below that threshold; otherwise its discrimination component remains unresolved. This does not require an automatic equivalence test at the outset.

H3 (conditional scope credit). Explicit hidden-agent content increases the recognition deficit: the number of unsupported premises independently coded in an account minus the number a participant identifies as unsupported, compared with a nonagent account with a matched coded premise count. Felt completeness is analyzed as a separate outcome. The deficit tests whether hidden causes obscure their added assumptions without comparing arbitrary rating scales.

H4 (production–reception asymmetry; exploratory). Effortful, analytic production of a detailed hidden account is separable from its reception, which depends on coherence, perceived effort, and identity relevance and not on repeating the producer's reasoning. This is a novel proposal with no direct evidence.

H5 (trust moderation). The H1 residual effect and the H2 completeness effect are predicted to be larger under a poor institutional competence history or lower measured trust; the latter moderation is exploratory in the separately recruited explanation cohort unless institutional history is also manipulated there.

Table 1

Evidence Relevant to the Premises of the Account

Premise	Status	Evidence
Hidden actors are readily generated as candidate explanations	Supported (correlational)	Agency detection, illusory pattern perception, and proportionality heuristics predict endorsement (Douglas et al., 2016; Leman & Cinnirella, 2007; van Prooijen et al., 2018).
Motives raise attraction to hidden accounts	Supported, small and heterogeneous	Epistemic, existential, and social motives (Biddlestone et al., 2025; Douglas et al., 2017); closure findings mixed (Leman & Cinnirella, 2013; Marchlewska et al., 2018).
Felt understanding exceeds demonstrated understanding	Supported in one panel study	Illusion of explanatory depth associated with endorsement (Vitriol & Marsh, 2018).
Endorsement reflects rejection of the official account more than coherence of an alternative	Supported (indirect)	Mutually contradictory theories endorsed together (Wood et al., 2012); anomaly-focused argument in one corpus (Wood & Douglas, 2013).

Premise	Status	Evidence
General complexity advantage	Contradicted or mixed	Preferences for simplicity and narrow scope; no credibility gain from added explanatory content; fact-based accounts judged broader (Khemlani et al., 2011; Lombrozo, 2007; Marsh & Kleinberg, 2025; Meuer et al., 2021).
High need for cognition among believers	Contradicted for believers; untested for producers	Weak negative association with need for cognition; analytic induction reduced endorsement; cognitive reflection correlates negatively (Bowes et al., 2023; Kantorowicz-Reznichenko et al., 2022; Swami et al., 2014).
Distrust moderates hidden-account appeal	Plausible, not established for the proposed interaction	Epistemic vigilance (Sperber et al., 2010); institutional error need not transfer trust (Dawson et al., 2025); corruption covaries with belief (Alper, 2023).
Updating on fraud beliefs reflects irrational asymmetry	Contradicted in one major case	A Bayesian belief-system model fit partisan changes better than irrational motivated updating (Botvinik-Nezer et al., 2023).
Concealment selectively down-weights absence evidence	Untested as specified	Content analyses show lower falsifiability but no matched belief updating (Meuer et al., 2023); self-sealing structure described conceptually (Boudry & Hofhuis, 2024).
Production–reception asymmetry	Untested	No study samples producers and receivers with independent measures.

Note. "Contradicted" means that the cited findings are inconsistent with the premise in the form tested. "Untested" means no study crosses the relevant features. Existing null findings generally did not cross concealment, source competence, and matched absence evidence, so they do not by themselves refute the conditional claims in H1.

Distinguishing the Account From Its Rivals

Several accounts can produce part or all of this pattern. Motivated reasoning can target identity-incongruent evidence (Taber & Lodge, 2006), and ambiguous absence may invite especially selective scrutiny; apparent bias can also reflect different warranted priors (Hahn & Harris, 2014). Source credibility predicts discounting evidence from a distrusted source (Sperber et al., 2010), whereas a rational belief-system account permits different updates when recipients use different causal models (Botvinik-Nezer et al., 2023). Self-sealing accounts describe concealment claims that insulate a belief system (Boudry & Hofhuis, 2024). The proposed experiment holds the immediate evidence source constant, elicits priors, and supplies item likelihoods; it can test excess weighting under those constraints. It cannot establish a unique psychological process if motivated reasoning also predicts selective treatment of missing records. Table 2 makes this overlap explicit.

Table 2

Predictions of Rival Accounts Under the Proposed Design

Account	Higher baseline belief in H	Discounts all institutional evidence	Concealment × evidence-type interaction	Excess update after absence at c = 1
Generic motivated reasoning	If identity-congruent	Yes, for incongruent evidence	Possible when missing evidence is more ambiguous; may depend on identity	Possible; not unique to this account
Source distrust	Under distrust	Yes	No (evidence source held constant)	Not predicted by source-only discounting with the same auditor
Rational belief systems	Via priors	Only via priors or causal models	Only through the stipulated or inferred *c*	Possible if inferred likelihoods differ; elicited benchmark tests this
Self-sealing structure	Not specified	Not specified	Qualitatively yes; no parameter or evidence-type prediction	Possible, without quantitative benchmark
Concealment weighting (this account)	Not required	No	Residual interaction under stated likelihoods	C = 1 residual exceeds neutral-item update

Note. All immediate test evidence is attributed to the same independent auditors, and likelihoods are specified prospectively. Under the symmetric-lookalike assumption, complete suppression makes a failed search neutral. A positive update may be rational under a participant's independently justified causal model; the personalized benchmark and neutral-item control address different alternatives.

The account's incremental claim is a quantified, prospectively benchmarked evidence-type interaction while prior belief and immediate source are controlled. It specifies when concealment should alter an absence-based update and how far an observed update exceeds that allowance. This can improve prediction without proving a distinct process: if motivated reasoning, source credibility, or an elicited causal model explains the same residuals equally well, the account has no demonstrated explanatory advantage. Identity relevance in Study 3 offers a further, observational comparison rather than a decisive mechanism test.

Constructs and Measurement

Table 3
Constructs and Proposed Measures

Construct	Measure and interpretation
Felt completeness	Rating of how many of the stated observations the account explains, collected before new test evidence; reported separately from accuracy.
Predictive constraint	Probability assigned before the evidence is revealed to each independently checkable trace under each account; scored against the generating process.
Hidden-cause gap	Joint pattern of completeness and discrimination, reported as two separate estimates and, if useful, their standardized association. Never a raw subtraction.
Exploratory sensitivity	Exploratory group-level slope and intercept of log-odds change on assigned log Bayes factor across items. Primary inference uses residuals.
Excess confirmation	Difference between residual updating after a $c = 1$ failed search and after a nondiagnostic no-concealment control (both $\ell = 0$); report condition means and elicited benchmark.
Evidence threshold	Number or diagnostic strength of matched items needed for a preregistered belief change, estimated by adaptive evidence sequences.
Discrimination	AUC or signal-detection sensitivity for generated H versus O cases; report the interval against a prespecified minimum worthwhile gain and response criterion separately.
Recognition deficit	Independently coded number of unsupported premises minus participant's count in the same units; compare actor-present with premise-count-matched nonagent accounts.

Note. Probability judgments use a 0–100 scale. Endpoint values remain in descriptive summaries and are mapped to prespecified finite logits only in the primary analysis. Endpoint rates, baseline beliefs, and alternative ordinal or bounded-response analyses are reported. Participants' estimates of P(find | H) and P(find | O) yield a personalized benchmark; this does not establish that their assumptions are correct. Premise recognition is measured as a count in the same units as independent coders' count.

Research Agenda

Table 4
Overview of Proposed Studies and Sample Planning

Study and design	Key outcomes	Hypotheses	Planned sample
1. Paired-account content analysis	Prospective concealment claims; coded absent traces	Context for H1	Event and text counts fixed by registry; precision target preregistered
2A. Fictional updating cohort	Residual update; neutral-controlled excess confirmation	H1; H5 competence	Pilot-informed crossed participant/scenario power simulation
2B. Explanation and discrimination cohort	Completeness; recognition deficit; AUC/criterion	H2; H3	Separate precision planning for minimum worthwhile AUC gain
3. Real-event hypotheticals	Residual updates by measured trust and prior	H5; H1 generalization	Independent precision and ethics plan after event selection
4. Experimental generation and reception	Writer effort and reader evaluation	H4 exploratory	Precision-based recruitment after stimulus pilot

Note. Numerical recruitment targets are intentionally deferred until the materials and variance components have been piloted. Study 2 uses separate cohorts. AUC = area under the receiver-operating characteristic curve.

Study 1: Content analysis of absence reports in paired accounts

Study 1 extends Meuer et al. (2023), who analyzed paired online articles and found more emotional and threat-related content and lower judged falsifiability in conspiratorial texts. Our aim is to test a different feature: how each account handles *missing* evidence. We sample events first, from a preregistered registry of incidents with both an institutional account and an alternative hidden-cause account, stratified by domain and region, and include events in which the alternative account was later verified as well as events in which it was not. Only then do we select texts, using date windows and source rules and not traction or persuasiveness.

The coding unit is an *absence report*: any statement that an expected trace was not found. Two trained coders, working from a preregistered manual, record for each unit whether an auxiliary explanation is offered (for example, records lost or suppressed), whether it is prospective or retrospective relative to the failed search, whether it specifies a mechanism, and whether it names an observable consequence that would count against it. Institutional accounts also use auxiliaries, so both account types are coded identically. The outcome is the *proportion* of absence reports with retrospective, unspecified auxiliaries, which avoids the trivial result that only alternative accounts mention concealment. Multilevel logistic models with events as random effects test the prediction that this proportion is higher in alternative accounts and lower in verified cases. Coder agreement is reported per construct after a 20-text pilot, with a preregistered threshold ($\kappa \geq .70$), and blinded raters score valence, arousal, readability, and source credibility. Because rhetoric and content co-vary, Study 1 is descriptive and cannot show an effect on persuasion.

Study 2: Controlled updating and explanation comparisons (tests of H1–H3)

Study 2 has two independently recruited cohorts to limit fatigue. The primary updating cohort sees 12 fictional incidents, each with an institutional and a hidden-cause account. Scenarios rotate across concealment and competence-history conditions between participants while the dossier and evidence source remain fixed. A separate explanation-and-discrimination cohort sees 16 fictional incidents sampled evenly from a larger pretested pool. In that cohort, an alternative account either identifies a covert actor and motive or acknowledges an unspecified gap while offering nonagent auxiliary explanations. Explicit concealment content is crossed independently with actor content. Each version preserves the stated observations and is matched for length, emotional tone, baseline plausibility, and the independently coded number of added unsupported premises. The pilot checks whether this matching succeeds; otherwise the H3 contrast is restricted or redesigned before preregistration.

In the updating cohort, concealment efficacy (none, partial $c = .5$, complete $c = 1$) and competence history (accurate/transparent versus error-prone/opaque) are crossed and counterbalanced. "None" means the account makes no concealment claim; it does not assert that concealment was impossible. Diagnosticity is stated separately in the test information, including $c = 0$ for this specified search, without adding a narrative denial. For $c = 0$ and $c = .5$, absence and direct-contradiction items have comparable negative log Bayes factors in the feasible common range. One anchor is BF $\approx .60$: at $c = 0$ set $tq = .40$, and at $c = .5$ set $tq = .80$. Other assigned factors stay above .5 and below 1. To match factors across c, t or q must differ, so those parameters are randomized across scenarios and their perceived plausibility and salience are checked in a blinded pilot. Item order is counterbalanced. Complete suppression and the no-concealment neutral control are reserved for the separate excess-confirmation contrast. Framing is piloted for typical initial beliefs between 30% and 70%, without excluding actual ratings outside that interval.

Diagnosticity is presented in natural frequencies (Gigerenzer & Hoffrage, 1995). For the BF $\approx .60$ anchor, a no-concealment item might say that 50 of 100 group actions produce a relevant record and 80 of 100 surviving records would be found ($tq = .40$). A partial-concealment item instead says all 100 actions produce a record, 80 of 100 surviving records would be found, and half of diagnostic records are removed ($tq = .80$, $c = .5$). In both versions, 2 of 100 cases under either account yield an incidental lookalike independently. The neutral control presents a procedurally similar new search result explicitly equally probable under H and O and contains no concealment language; its wording and initial beliefs are piloted against the complete-suppression item. Participants rate belief before and after an item and estimate P(find | H) and P(find | O) after the frequency description but before the result. Stipulated and elicited log Bayes

factors are computed separately. A pilot checks comprehension, demand, plausibility, salience, and whether differing t or q creates a detectable confound.

In the explanation-and-discrimination cohort, each participant receives 16 scenarios balanced across actor content, concealment claim, and generating state (eight H and eight O); scenario-to-cell assignment rotates between participants. After reading an account but before test evidence, participants rate felt completeness, count unsupported premises, and forecast a held-out trace. Independent coders count added unsupported premises using a preregistered manual, permitting a same-unit recognition deficit. The subsequent evidence is generated under the stated H or O process and includes both found and absent records and an independently sourced direct item; it is not conditioned to show a failed search. Participants then rate final belief and confidence. Truth labels and final beliefs yield discrimination sensitivity and response criterion, with crossed scenario effects (Robson et al., 2026).

Primary estimand and sample size. In the updating cohort, the primary effect is the concealment × evidence-type interaction in residual updating $r = u - \ell$ for negative items, with crossed participant and scenario effects. The complete-suppression versus neutral-control contrast is a distinct planned test. Personalized-benchmark residuals, trust moderation, completeness, recognition deficits, held-out forecasts, and discrimination are separately reported outcomes; the slope-and-intercept model is exploratory. A planning difference of 0.10 log-odds units per item requires pilot justification. Power simulations will use actual assignment, participant and scenario variance, item clustering, endpoint rates, and missingness. Each cohort has its own preregistered sample-size calculation and inferential thresholds; neither cohort inherits power from the other (Westfall et al., 2014).

Interpretation. Support for the H1 primary effect requires a confidence interval in the predicted direction and a point estimate exceeding a preregistered meaningful threshold; intervals inside prespecified equivalence bounds count against a meaningful effect, using two one-sided tests where appropriate (Lakens, 2017). The neutral-controlled excess-confirmation test is separate. H2 requires a positive completeness contrast; the discrimination difference and its interval are reported against a preregistered minimum worthwhile gain. Only an interval excluding a gain of that size warrants a “limited discrimination gain” conclusion; a nonsignificant result is inconclusive. H3 requires a positive between-account recognition-deficit contrast after independently coded premise counts are matched. Multiplicity treatment and decision rules are registered before confirmatory data collection.

Study 3: Real-event hypotheticals with measured distrust (H5; generalization of H1)

Fictional cases lack the identity stakes and histories surrounding real events. Study 3 uses neutral summaries of at least four real events, including documented covert coordination and

unsupported allegations, selected across political and cultural contexts. Participants consider explicitly hypothetical future evidence with stated diagnosticity; invented observations are never presented as facts. Trust, prior belief, and identity relevance are measured before the task. The protocol will receive ethics approval, use a consent warning for politically charged material, and end with debriefing that clearly separates documented history from hypothetical evidence. Baseline probabilities near 0 or 100 are retained; preregistered nonlinear splines and a 20%–80% sensitivity subset address compression. Endpoint rates and bounded-response models are reported. Measured distrust can moderate associations but does not establish a causal effect of trust.

Study 4: Production and reception (exploratory test of H4)

Study 4 recruits ordinary participants as *experimental generators*, not presumed originators of naturally circulating theories. Each reads the same ambiguous dossier and is randomly assigned to write a conventional explanation, a covert-coordination explanation, or their best account. Need for cognition and analytic thinking are measured before writing. Preregistered coding assesses the number of independently testable predictions, hidden assumptions, causal links, and post hoc immunizing moves, controlling for length and fluency. A separate audience sample, blind to generator traits, rates anonymized, length-matched accounts under ordinary evaluation or an evidence audit that requires the same causal reconstruction and falsifier for each account. This crossed design separates effortful production from persuasive reception. Generalization to organic originators would require a separate observational study, and failure to find the asymmetry would not by itself defeat H1.

Limitations and Conclusion

This theory paper uses a targeted, nonexhaustive evidence map and makes no systematic-review claim. The central interaction remains untested; broad claims about complexity or believers' unusually high need for cognition are unsupported. The search benchmark depends on assumptions about trace generation, suppression, detection, and equal incidental-lookalike rates. Participants may reasonably reject the stipulated rates, so stipulated and elicited benchmarks are reported; the neutral control addresses a general upward response tendency. Matching Bayes factors requires different trace-production or search rates across conditions and is limited to their common feasible range. Even with piloting, these changed rates may introduce other differences. Equivalence-grade precision for discrimination may require far more participants than the updating test, leaving H2's second component unresolved. The model does not establish how hidden accounts originate or spread.

The conditional prediction is that alleged concealment selectively weakens updating after a failed search more than its specified efficacy warrants, while equally diagnostic direct

contradictions retain their force. Under complete suppression, absence is uninformative under the symmetric-lookalike assumption; a neutral-item control distinguishes selective excess confirmation from general response tendencies. An explicit hidden actor may increase felt completeness and reduce recognition of unsupported premises, while its effect on discrimination requires a separately precise estimate. Controlled null results would require rejecting or narrowing these claims.

References

Alper, S. (2023). There are higher levels of conspiracy beliefs in more corrupt countries. *European Journal of Social Psychology, 53*(3), 503–517. https://doi.org/10.1002/ejsp.2919

Biddlestone, M., Green, R., Douglas, K. M., Azevedo, F., Sutton, R. M., & Cichocka, A. (2025). Reasons to believe: A systematic review and meta-analytic synthesis of the motives associated with conspiracy beliefs. *Psychological Bulletin, 151*(1), 48–87. https://doi.org/10.1037/bul0000463

Botvinik-Nezer, R., Jones, M., & Wager, T. D. (2023). A belief systems analysis of fraud beliefs following the 2020 US election. *Nature Human Behaviour, 7*, 1106–1119. https://doi.org/10.1038/s41562-023-01570-4

Boudry, M., & Hofhuis, S. (2024). On epistemic black holes: How self-sealing belief systems develop and evolve. *Theoria, 90*(4), 429–447. https://doi.org/10.1111/theo.12554

Bowes, S. M., Costello, T. H., & Tasimi, A. (2023). The conspiratorial mind: A meta-analytic review of motivational and personological correlates. *Psychological Bulletin, 149*(5–6), 259–293. https://doi.org/10.1037/bul0000392

Dawson, A. J., Bista, A., & Wilson, A. E. (2025). Is trust a zero-sum game? What happens when institutional sources get it wrong. *PLOS ONE, 20*(4), Article e0321743. https://doi.org/10.1371/journal.pone.0321743

Dentith, M. R. X. (2014). *The philosophy of conspiracy theories*. Palgrave Macmillan.

Douglas, K. M., Sutton, R. M., Callan, M. J., Dawtry, R. J., & Harvey, A. J. (2016). Someone is pulling the strings: Hypersensitive agency detection and belief in conspiracy theories. *Thinking & Reasoning, 22*(1), 57–77. https://doi.org/10.1080/13546783.2015.1051586

Douglas, K. M., Sutton, R. M., & Cichocka, A. (2017). The psychology of conspiracy theories. *Current Directions in Psychological Science, 26*(6), 538–542. https://doi.org/10.1177/0963721417718261

Douglas, K. M., & Sutton, R. M. (2023). What are conspiracy theories? A definitional approach to their correlates, consequences, and communication. *Annual Review of Psychology, 74*, 271–298. https://doi.org/10.1146/annurev-psych-032420-031329

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review, 102*(4), 684–704. https://doi.org/10.1037/0033-295X.102.4.684

Hahn, U., & Harris, A. J. L. (2014). What does it mean to be biased: Motivated reasoning and rationality. In B. H. Ross (Ed.), Psychology of learning and motivation (Vol. 61, pp. 41–102). Academic Press. https://doi.org/10.1016/B978-0-12-800283-4.00002-2

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review, 114*(3), 704–732. https://doi.org/10.1037/0033-295X.114.3.704

Hornsey, M. J., Bierwiaczonek, K., Sassenberg, K., & Douglas, K. M. (2023). Individual, intergroup and nation-level influences on belief in conspiracy theories. *Nature Reviews Psychology, 2*, 85–97. https://doi.org/10.1038/s44159-022-00133-0

Kantorowicz-Reznichenko, E., Reinders Folmer, C., & Kantorowicz, J. (2022). Don't believe it! A global perspective on cognitive reflection and conspiracy theories about COVID-19 pandemic.

Personality and Individual Differences, 194, Article 111666. https://doi.org/10.1016/j.paid.2022.111666

Khemlani, S. S., Sussman, A. B., & Oppenheimer, D. M. (2011). Harry Potter and the sorcerer's scope: Latent scope biases in explanatory reasoning. *Memory & Cognition, 39*(3), 527–535. https://doi.org/10.3758/s13421-010-0028-1

Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science, 8*(4), 355–362. https://doi.org/10.1177/1948550617697177

Lantian, A., Muller, D., Nurra, C., & Douglas, K. M. (2017). "I know things they don't know!" The role of need for uniqueness in belief in conspiracy theories. *Social Psychology, 48*(3), 160–173. https://doi.org/10.1027/1864-9335/a000306

Leman, P. J., & Cinnirella, M. (2007). A major event has a major cause: Evidence for the role of heuristics in reasoning about conspiracy theories. *Social Psychological Review, 9*(2), 18–28.

Leman, P. J., & Cinnirella, M. (2013). Beliefs in conspiracy theories and the need for cognitive closure. *Frontiers in Psychology, 4*, Article 378. https://doi.org/10.3389/fpsyg.2013.00378

Lombrozo, T. (2007). Simplicity and probability in causal explanation. *Cognitive Psychology, 55*(3), 232–257. https://doi.org/10.1016/j.cogpsych.2006.09.006

Marchlewska, M., Cichocka, A., & Kossowska, M. (2018). Addicted to answers: Need for cognitive closure and the endorsement of conspiracy beliefs. *European Journal of Social Psychology, 48*(2), 109–117. https://doi.org/10.1002/ejsp.2308

Marsh, J. K., & Kleinberg, S. (2025). Go big or go hoax: Explanatory scope and the believability of conspiracy theories. Proceedings of the Annual Meeting of the Cognitive Science Society, 47. https://escholarship.org/uc/item/7gc219mf

Meuer, M., Oeberst, A., & Imhoff, R. (2021). Believe it or not: No support for an effect of providing explanatory or threat-related information on conspiracy theories' credibility. *International Review of Social Psychology, 34*(1), Article 26. https://doi.org/10.5334/irsp.587

Meuer, M., Oeberst, A., & Imhoff, R. (2023). How do conspiratorial explanations differ from non-conspiratorial explanations? A content analysis of real-world online articles. *European Journal of Social Psychology, 53*(2), 288–306. https://doi.org/10.1002/ejsp.2903

Pytlik, N., Soll, D., & Mehl, S. (2020). Thinking preferences and conspiracy belief: Intuitive thinking and the jumping to conclusions bias as a basis for the belief in conspiracy theories. *Frontiers in Psychiatry, 11*, Article 568942. https://doi.org/10.3389/fpsyt.2020.568942

Robson, S. G., Martire, K. A., Pearson, S., Cosgrove, T. J., Jetten, J., Faasse, K., & Hornsey, M. J. (2026). Rejecting conspiracies doesn't necessarily mean that you're smart: The relationship between conspiracy discrimination and reasoning ability. *British Journal of Social Psychology, 65*(4), Article e70117. https://doi.org/10.1111/bjso.70117

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language, 25*(4), 359–393. https://doi.org/10.1111/j.1468-0017.2010.01394.x

Swami, V., Voracek, M., Stieger, S., Tran, U. S., & Furnham, A. (2014). Analytic thinking reduces belief in conspiracy theories. *Cognition, 133*(3), 572–585. https://doi.org/10.1016/j.cognition.2014.08.006

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science, 50*(3), 755–769. https://doi.org/10.1111/j.1540-5907.2006.00214.x

van Prooijen, J.-W., Douglas, K. M., & De Inocencio, C. (2018). Connecting the dots: Illusory pattern perception predicts belief in conspiracies and the supernatural. *European Journal of Social Psychology, 48*(3), 320–335. https://doi.org/10.1002/ejsp.2331

Vitriol, J. A., & Marsh, J. K. (2018). The illusion of explanatory depth and endorsement of conspiracy beliefs. *European Journal of Social Psychology, 48*(7), 955–969. https://doi.org/10.1002/ejsp.2504

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General, 143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Wood, M. J., & Douglas, K. M. (2013). "What about building 7?" A social psychological study of online discussion of 9/11 conspiracy theories. *Frontiers in Psychology, 4*, Article 409. https://doi.org/10.3389/fpsyg.2013.00409

Wood, M. J., Douglas, K. M., & Sutton, R. M. (2012). Dead and alive: Beliefs in contradictory conspiracy theories. *Social Psychological and Personality Science, 3*(6), 767–773. https://doi.org/10.1177/1948550611434786

Appendix: Correlated Traces and Alternative Incidental Rates

When searches for a primary record and a secondary sign of suppression are dependent, their joint likelihood should be specified under both accounts. Multiplying separate Bayes factors would count shared information twice. If incidental lookalikes occur at different rates under H and O, the failed-search factor becomes $(1 - \tau)(1 - f_H)/(1 - f_O)$, under conditional independence of incidental and diagnostic records. At complete suppression, its value can depart from one; that departure reflects justified asymmetric incidental rates, not the absence of the diagnostic record itself. The proposed experiment holds the incidental rate equal under both accounts.