

**The Attribution Problem:
Causal and Informational Value in Derivative Media**

PIATRA . INSTITUTE

October 2026

Abstract

A derived object, such as a reaction video, an explainer, a remix or a summary, takes something from a source and adds something of its own, and its value is usually reported as one number. We separate four quantities the number collapses. Inherited, independent and interaction value are identities of one value function, identified by randomizing viewers over four exposures; the externality on the source is a different causal quantity that needs a different design. Under Blackwell's order a derivative that is a garbling of its source cannot raise a Bayesian viewer's value. Attention reverses this only when its price attaches to the input stream, and in a needle-in-a-stream model the reversal has an exact value: a reaction of 25 segments to a source of 90 is worth up to 0.357 more, and between prices of 0.0110 and 0.0385 per segment only the reaction is worth scanning. For a calibrated viewer the epistemic gain under a Brier score is the variance of the belief update, and a signal of accuracy 0.7 stops paying once the reported confidence passes 0.90. Shapley allocations are not identified by any experiment; the data identify Harsanyi dividends. In a 16-cell design for four reaction modules a Shapley share needs 2,133 observations for a standard error of 0.05 standard deviations, a pair dividend 25,600 and the four-way dividend 102,400. A parametric-bootstrap test of the garbling bound has size 0.08 to 0.11 at the nominal 0.05 and 0.03 to 0.06 at the nominal 0.025. Three hundred videos in a field experiment detect only effects of about 28% on views, within-creator randomization misses catalogue spillover, and creator-level randomization of 30 creators detects only 49%. A sheaf consistency energy is a heterogeneity statistic whose blind spots are set by its cover: a marginal cover cannot see a pure interaction, with power 0.05 against 0.95 for the homogeneity test.

1. The attribution problem

A derived object takes something from a source and adds something of its own. A reaction video shows part of a source and talks over it, an explainer retells a paper, a remix reuses a song, a fork reuses software, a summary compresses a book. What the derived object is worth, and to whom, is asked in copyright law as a question of transformation and market substitution and in platform economics as a question of revenue shares, and both ask for one number. The quantities involved are at least four: what the source already supplied, what the derivative supplies alone, what exists only in the combination, and what happens to the source and to others because the derivative exists. They have different causal definitions, different designs that identify them and different costs in observations.

Reaction videos are the first instance because the source can be observed beside its derivative. Scholarship on them studies why viewers watch, through the comments under four reaction channels (Ghosh and Tripathi 2025), and how fair use treats audiovisual critique (Fisher 2021). Neither measures what a reaction adds to a viewer or takes from a creator. We do not take up the legal question: the quantities below are empirical, and whether any of them bears on a legal test is outside the paper.

The tools are known. Blackwell's order compares experiments, rational inattention prices attention, proper scores measure belief, the Shapley value and the Harsanyi dividends divide a game, partial information decomposition splits a mutual information, and interventional versions of the Shapley value use the do-operator (Blackwell 1953; Sims 2003; Gneiting and Raftery 2007; Shapley 1953; Harsanyi 1959; Williams and Beer 2010; Janzing et al. 2019; Witter et al. 2026). The literature reviewed for this paper contains no treatment of derivative media with them; that is a statement about that literature and not a claim of priority. What the paper adds is their assembly into one measurement program, the exact calculations where an exact calculation exists, and the cost in observations of each proposed estimator.

Inherited, independent and interaction value are identities of one value function and are identified by randomization; the interaction is reported before any split of it. A derivative that is a garbling of its source cannot add value for an unconstrained Bayesian viewer, and attention-mediated value requires a price on the input. Epistemic gain under a proper score decomposes into resolution and miscalibration. The empirical test of the garbling bound is a boundary problem with a sample-size-dependent null. Shapley allocations are not identified, partial information decomposition has a finite-sample bias that sets its sample size, and the 16-cell design resolves shares and main effects at feasible sample sizes but not high-order dividends. Field designs identify the source externality only up to catalogue spillover and with limited power. A sheaf consistency energy detects only what its cover lets it detect.

2. Four quantities

Let C be a source object and R a reaction layer, an operation that selects, compresses, comments on and reframes the source. The derived object, written CR , is the result of applying R to C . For a viewer i with context x , let $Y_i(\omega)$ be the outcome vector under exposure $\omega \in \{\emptyset, C, R, CR\}$: nothing, the source alone, the reaction layer without the source footage, and the derived object. The viewer-side outcomes are epistemic value E , hedonic value H , social and affective value S , attention efficiency A and discovery of the source or its creator D . A further family belongs to a different unit: the outcomes of the source, such as its views, watch time and revenue, gathered in M .

The primitive quantity is a vector:

$$\Delta_R(x) = \mathbb{E}[\mathbf{Y} \mid do(CR), x] - \mathbb{E}[\mathbf{Y} \mid do(C), x].$$

A weight vector w turns it into a value, $v_w(\omega) = w^\top \mathbb{E}[\mathbf{Y}(\omega) \mid x]$. Choosing w states how much enjoyment weighs against learning or against creator revenue and is a normative act. Estimating $\mathbf{Y}(\omega)$ is empirical and does not depend on w .

Four quantities are usually collapsed. Inherited value is $v(C) - v(\emptyset)$, what the source supplies. Independent value is $v(R) - v(\emptyset)$, what the reaction layer supplies alone. Interaction value is the Möbius coefficient

$$\Gamma_{CR} = v(CR) - v(C) - v(R) + v(\emptyset),$$

what exists only in the combination. Externality is the effect of the existence of CR on outcomes elsewhere, $\Delta M = \mathbb{E}[M \mid CR \text{ exists}] - \mathbb{E}[M \mid CR \text{ absent}]$. The first three are identities of one value function:

$$v(CR) - v(\emptyset) = [v(C) - v(\emptyset)] + [v(R) - v(\emptyset)] + \Gamma_{CR},$$

so the contribution of the reactor given the source is its independent value plus the interaction, $v(CR) - v(C) = [v(R) - v(\emptyset)] + \Gamma_{CR}$, and the contribution of the source given the reactor is its inherited value plus the same interaction. The two-player Shapley value gives the reactor $[v(R) - v(\emptyset)] + \Gamma_{CR}/2$ and the source $[v(C) - v(\emptyset)] + \Gamma_{CR}/2$: it splits the interaction in half whatever its size. Γ_{CR} is an observable quantity and should be reported before any split.

Randomizing viewers over the four exposures identifies every $\mathbb{E}[Y(\omega) \mid x]$, hence Γ_{CR} and the inherited and independent values, for any w . The externality needs units on the source's side and a design in which the existence of the reaction varies independently of the source's characteristics. The two are different causal quantities: value created for viewers, and value displaced from or promoted to creators. A reaction can create value for every viewer while displacing the creator's own audience, and the sign of one says nothing about the sign of the other.

3. Information and decision theory

3.1 The garbling bound

Treat an exposure as a statistical experiment: a row-stochastic matrix κ from the state of the world $\theta \in \Theta$ to the viewer's resulting signal z .

Proposition 1. If $\kappa_{CR} = \kappa_C G$ for a row-stochastic matrix G , then for every finite decision problem $(\mathcal{A}, u, \pi)$ a Bayesian viewer without processing costs obtains at most as high an expected utility from CR as from C .

Proof. Let δ be any decision rule on the signals of CR . Define the rule $\delta'(z) = \sum_{z'} G(z, z') \delta(z')$ on the signals of C . Its expected utility under κ_C equals that of δ under $\kappa_C G = \kappa_{CR}$. Hence the optimal value under C is at least that under CR . $\square$

Blackwell's theorem is the converse: if C is worth at least as much as CR in every decision problem, a garbling exists (Blackwell 1953). When $V(CR) > V(C)$ for a viewer, at least one hypothesis of Proposition 1 fails: CR is not a garbling of C and adds information, the viewer faces processing costs, or utility depends on something besides the action and the state, namely enjoyment, social value or the frame in which the information arrives (Tversky and Kahneman 1981). A Monte Carlo check on 20,000 random experiments, garblings and decision problems found no violation beyond rounding error.

3.2 When attention reverses the ranking

Rational inattention charges the viewer for information, and the form of the charge decides the answer. Costs defined on the information structure the viewer ends up with, such as the Shannon and the posterior-separable costs, are the standard ones, and they give choice rules such as the generalized multinomial logit (Sims 2003; Matějka and McKay 2015; Caplin and Dean 2015; Caplin et al. 2022; Maćkowiak et al. 2023).

Proposition 2. (a) If the viewer chooses among the information structures available from an exposure and pays a cost that depends only on the structure chosen, and $\kappa_{CR} = \kappa_C G$, then $V_\lambda(C) \geq V_\lambda(CR)$ for every attention price λ . (b) If the price attaches to the input stream, there are exposures with $\kappa_{CR} = \kappa_C G$ and $V_\lambda(CR) > V_\lambda(C)$.

Proof. (a) The structures available from an exposure are the garblings of its kernel. A garbling of $\kappa_{CR} = \kappa_C G$ is a garbling of κ_C , so the feasible set of CR is contained in that of C and the same cost function is maximized over a smaller set. (b) The needle model below. $\square$

A viewer who maximizes expected utility minus a cost on her attention policy therefore cannot prefer the garbling when the cost is a function of the information she acquires, since she can replicate the garbling from the richer source. The familiar case of a reactor who turns a difficult 90-minute video into an understandable 25-minute one needs the price to attach to the stream.

The needle model. The state $\theta \in \{0, 1\}$ is equally likely. The source is a stream of T segments, one of which, at a uniformly random position, reveals θ ; the others are noise. Processing a segment costs λ in utility. The viewer then chooses an action and receives 1 if it matches θ . After scanning k segments without finding the needle her belief is unchanged, and the value of scanning k segments and then guessing is $\frac{1}{2} + \frac{k}{2T} - \lambda \left[\frac{k(k+1)}{2T} + \frac{k(T-k)}{T} \right]$, which is convex in k . The optimum is therefore at $k = 0$ or $k = T$:

$$V_\lambda(T) = \max \left\{ \frac{1}{2}, 1 - \frac{\lambda(T+1)}{2} \right\}.$$

The reactor sees the whole stream, finds the needle and replays it in a derived stream of $T' < T$ segments. The derived stream is a function of the source stream, so its kernel is a garbling of the source's. Its value is $V_\lambda(T')$. The gain $V_\lambda(T') - V_\lambda(T)$ equals $\lambda(T - T')/2$ for $\lambda \leq 1/(T + 1)$, equals $\frac{1}{2} - \lambda(T' + 1)/2$ for $1/(T + 1) < \lambda \leq 1/(T' + 1)$, and is zero beyond. It is non-decreasing in the source length T , strictly increasing until the viewer abandons the source at $T = 1/\lambda - 1$ and constant afterwards. This is the content of attention-mediated value: it needs a price on the input and a derived object that is cheaper to process than the source while carrying the same information about the state.

For $T = 90$ and $T' = 25$ the source is scanned only at prices up to $1/91 = 0.0110$ and the reaction up to $1/26 = 0.0385$. The gain peaks at $65/182 = 0.357$ at the price $1/91$ and falls to zero at $1/26$ (Table 1, Figure 1). Backward induction over the number of segments scanned reproduces the

closed form to machine precision, and scanning an intermediate number of segments never beats the better of $k = 0$ and $k = T$.

Table 1: Value of a source of 90 segments and of a reaction of 25 segments carrying the same information, by price of attention per segment.

price λ	value of the source	value of the reaction	gain
0.002	0.909	0.974	0.065
0.005	0.773	0.935	0.163
0.01	0.545	0.870	0.325
0.02	0.500	0.740	0.240
0.03	0.500	0.610	0.110
0.05	0.500	0.500	0.000

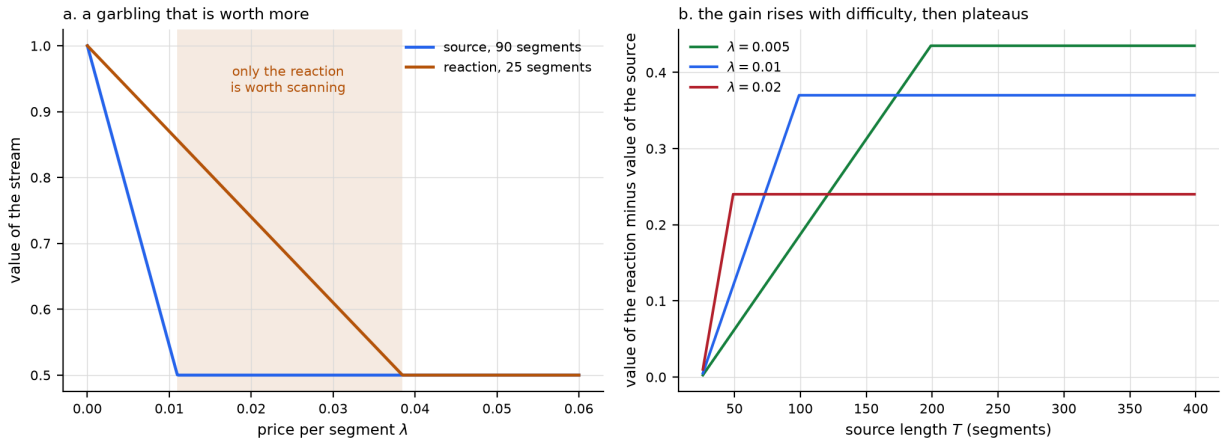


Figure 1: The needle model. (a) Value of the source stream and of the reaction stream against the price per segment; between the two thresholds only the reaction is worth scanning. (b) Value of the reaction minus value of the source against the source length, at three prices: the gain rises with the difficulty of the source and then stays constant once the viewer abandons the source.

3.3 Epistemic gain under a proper score

Epistemic value should be measured by prediction; a self-rated sense of having learned is a different quantity. Before exposure the viewer states a probability q_0 that an event $y \in \{0, 1\}$ occurs, after exposure q_1 , and each is scored by the Brier score $(q - y)^2$, a strictly proper rule (Brier 1950; Gneiting and Raftery 2007). The epistemic gain is $\Delta E = BS_{pre} - BS_{post}$.

Proposition 3. Let $p_0 = P(y = 1 \mid \mathcal{F}_0)$ and $p_1 = P(y = 1 \mid \mathcal{F}_1)$ be the Bayesian beliefs of a viewer before and after exposure, $\mathcal{F}_0 \subset \mathcal{F}_1$, and let q_0 be $\mathcal{F}_0$ -measurable and q_1 be $\mathcal{F}_1$ -measurable reports. Then

$$\mathbb{E}(q_0 - y)^2 - \mathbb{E}(q_1 - y)^2 = \mathbb{E}(p_1 - p_0)^2 + \mathbb{E}(q_0 - p_0)^2 - \mathbb{E}(q_1 - p_1)^2.$$

Proof. For an $\mathcal{F}_t$ -measurable q , $\mathbb{E}(q - y)^2 = \mathbb{E}(p_t - y)^2 + \mathbb{E}(q - p_t)^2$, because $\mathbb{E}[(p_t - y)(q - p_t)] = 0$. For the Bayesian beliefs, $\mathbb{E}(p_0 - y)^2 - \mathbb{E}(p_1 - y)^2 = \mathbb{E}(p_1 - p_0)^2$ because $\mathbb{E}[(p_1 - y)(p_1 - p_0)] = 0$. Combining the two gives the identity. $\square$

For a calibrated viewer the expected Brier gain is the variance of her belief update, the resolution term of the Brier decomposition (Murphy 1973), a term closely related to the refinement used to compare forecasters (DeGroot and Fienberg 1983). The other two terms measure miscalibration, and a reactor who persuades without informing, by pushing reported beliefs away from the Bayesian ones, makes the last term positive and the gain smaller or negative. The identity holds to rounding error on 2,000 random finite probability spaces with arbitrary reports. For a symmetric binary signal of accuracy a about an event of prior probability one half, reported with confidence p in the signalled state, the gain is $\frac{1}{4} - a + 2ap - p^2$, which equals $(a - \frac{1}{2})^2$ at $p = a$ and falls to zero at $p = 2a - \frac{1}{2}$ when $a < \frac{3}{4}$. At $a = 0.7$ the calibrated gain is 0.040 and the break-even confidence is 0.90, a log-odds scaling of 2.59 of the calibrated report; the log score punishes overconfidence harder, with break-even at 0.86. At $a = 0.8$ the Brier gain stays positive even for a report of certainty, and the log score breaks even at 0.96 (Figure 2).

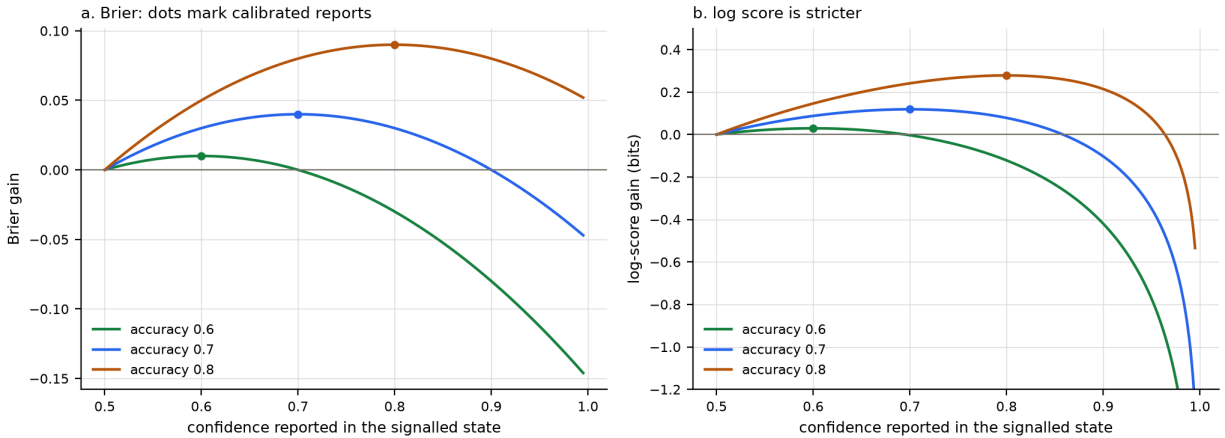


Figure 2: Gain from a signal of known accuracy against the confidence reported in the signalled state. (a) Brier gain; dots mark calibrated reports. (b) Log-score gain in bits.

3.4 Testing the garbling bound

Estimate the response channels $K_C(\theta, z)$ and $K_{CR}(\theta, z')$ from viewers' reports in each state and condition, and measure how far the derived channel is from the set of garblings of the source channel:

$$d_B = \min_G \sum_{\theta} \pi_{\theta} \text{KL}(K_{CR}(\theta, \cdot) \| (K_C G)(\theta, \cdot)),$$

with G over row-stochastic matrices and KL the Kullback-Leibler divergence (Cover and Thomas 2006). The objective is convex in G , and the minimum is found by the expectation-maximization algorithm (Dempster et al. 1977) with extrapolation to speed convergence; its output satisfies the optimality conditions to 2×10^{-16} and is never above a general-purpose optimizer. Here $d_B = 0$ if and only if the derived responses are a garbling of the source responses.

The plug-in estimate is positive under the null because the estimated channels are noisy, and the null sits on the boundary of the set of garblings, so the usual chi-square limit does not apply (Self and Liang 1987). We test with a parametric bootstrap that draws both samples from the source channel at its estimate and the derived channel at the fitted garbling (Efron and Tibshirani 1993). In the simulation the source channel has two states and three report levels, with probabilities 0.5, 0.3 and 0.2 in one state and the reverse in the other, and the derived object adds a binary signal of accuracy b , so that its alphabet has six symbols; $b = 0.5$ is the null. The population distance is 0.0056 nats at $b = 0.6$, 0.0223 at 0.7, 0.0664 at 0.8 and 0.174 at 0.9.

Under the null the mean plug-in distance is 0.0078 nats at 100 responses per state and condition, 0.0025 at 300, 0.00075 at 1,000 and 0.00027 at 3,000. At 100 responses the noise alone exceeds the population distance of a reactor who adds a signal of accuracy 0.6, so a fixed threshold on the plug-in value cannot separate a garbling from a mild addition. The bootstrap test is anti-conservative: with 1,000 replicates per cell its size at the nominal 0.05 is 0.105, 0.084, 0.075 and 0.093 at 100, 300, 1,000 and 3,000 responses, at the nominal 0.025 it is 0.056, 0.042, 0.028 and 0.041, and at the nominal 0.01 it is 0.018, 0.013, 0.010 and 0.014. We recommend the nominal 0.025 for a test of size near 0.05. At that level power against a reactor adding a signal of accuracy 0.6 is 0.16 at 100 responses, 0.42 at 300 and 0.92 at 1,000; against 0.7 it is 0.47, 0.95 and 1.00; against 0.8 it is 0.93, 1.00 and 1.00 (Figure 3). We did not derive the finite-sample law of the statistic; the excess size is the kind of distortion that a null on the boundary of the model produces.

4. Interaction and allocation

With n modules of the reaction layer present or absent, the value function v is defined on the 2^n subsets. The Harsanyi dividends (Harsanyi 1959) are the Möbius transform of v (Rota 1964; Grabisch 2016),

$$d_T = \sum_{S \subseteq T} (-1)^{|T|-|S|} v(S), \quad v(S) = \sum_{T \subseteq S} d_T,$$

and the Shapley value (Shapley 1953) is the sum of each dividend shared equally among its members, $\phi_j = \sum_{T \ni j} d_T / |T|$. A dividend d_T is the value that exists only when all modules of T are present. Shapley values attribute a model's prediction to its inputs (Lundberg and Lee 2017), and interventional versions of them use the value function $v(S) = \mathbb{E}[Y \mid do(S)]$ (Janzing et al. 2019; Heskes et al. 2020; Witter et al. 2026); in a randomized experiment that function is measured directly in every cell, and the do-Shapley value is the Shapley value of the measured game.

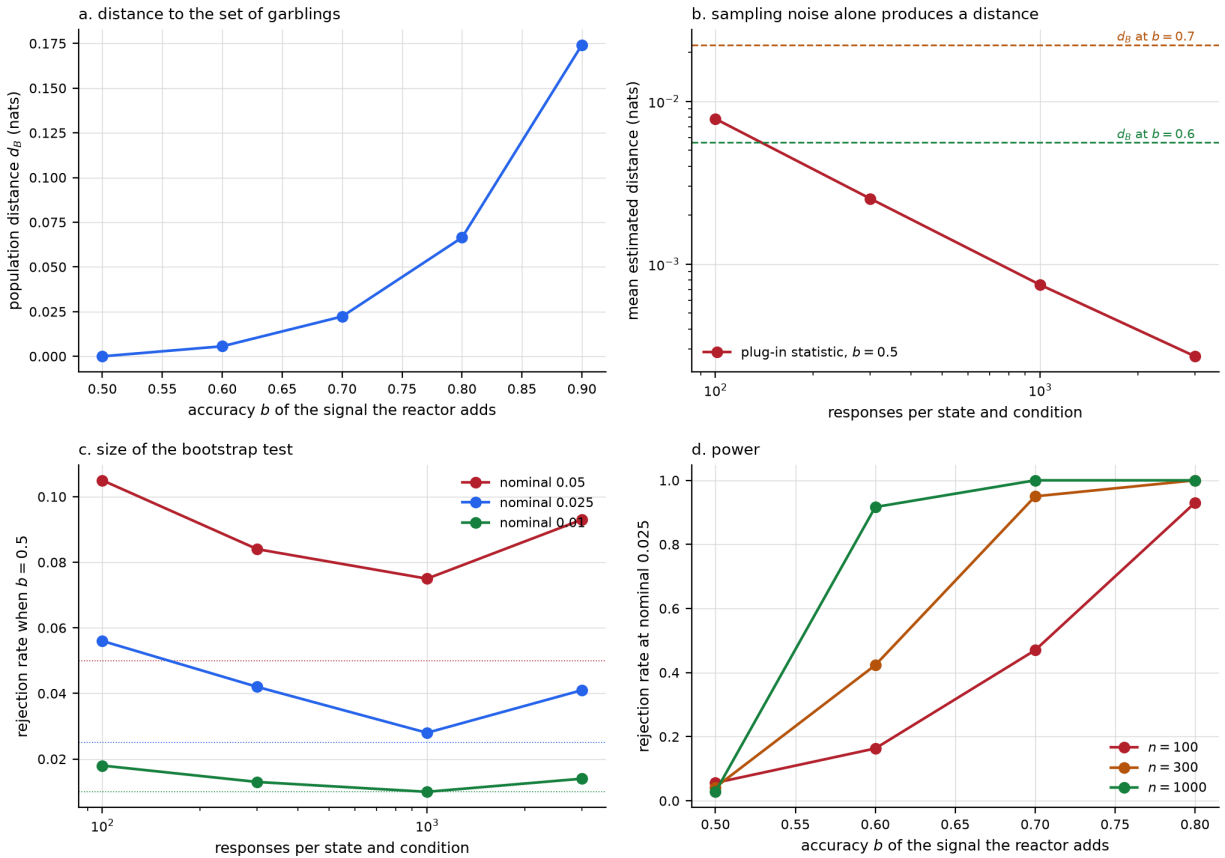


Figure 3: Empirical Blackwell test. (a) Population distance to the set of garblings against the accuracy of the signal the reactor adds. (b) Mean plug-in distance under the null against the number of responses; dashed lines are the population distances at accuracy 0.6 and 0.7. (c) Rejection rate under the null at three nominal levels. (d) Power at the nominal 0.025.

Proposition 4. The cell means identify the dividends. Any rule that assigns each dividend d_T to the modules of T in shares summing to one is consistent with the same data, and different rules give different allocations.

Proof. The dividends are linear functions of the cell means, so they are identified when the cell means are. The allocation rule is not a function of the data beyond the dividends; Shapley takes shares $1/|T|$ and a rule proportional to the stand-alone effects takes other shares. $\square$

The Shapley allocation therefore adds a rule to the measurement. Take a scenario with four modules, source selection and compression Q , interpretive commentary I , affective reaction A and editorial operations E , and the dividends of Figure 5a: main effects of 0.10, 0.30, 0.45 and 0.05 standard deviations for Q , I , A and E , a pair dividend of 0.15 for Q and I , of -0.10 for I and A and of 0.05 for I and E , and a triple dividend of 0.05 for Q , I and A . Their sum, the value of all modules together, is 1.05. The Shapley shares are 0.192 for Q , 0.367 for I , 0.417 for A and 0.075 for E . Adding each module's main effect to its share of the Q , I dividend gives 0.175 for Q and 0.375 for I under an equal split and 0.138 and 0.413 under a split in proportion to their stand-alone effects: the same data, a different allocation. Reporting the dividends leaves the allocation to the reader.

The same logic applies to a split of a reactor into information and persona by a neutral-information arm. Let U_0 be the source alone, U_N the source with the same proposition delivered neutrally, and U_R the source with the reactor delivering it. The contrasts $U_N - U_0$ and $U_R - U_N$ add to $U_R - U_0$ by construction, and read as information and persona they are one of two orders. With a fourth arm U_P , persona without the proposition, the other order gives $U_P - U_0$ and $U_R - U_P$, and the two orders differ by the interaction $(U_R - U_N) - (U_P - U_0) = d_{IA}$, the dividend of the information and persona modules. The decomposition is clean only if $d_{IA} = 0$, and the fourth arm is what shows whether it is.

5. Information decomposition

For a viewer outcome Y and two content variables, the source content C and the reaction content R , the mutual information splits into four parts, $I(C, R; Y) = Red + Unq_C + Unq_R + Syn$. In the reaction setting these are the repetition of what the source already carries (a reactor who says “yes, that is right”), a fact only the source supplies, a fact only the reactor supplies, and information that exists only in the juxtaposition, as when commentary makes a source intelligible. The split is not unique. The redundancy measure of Williams and Beer compares the amounts of information the sources carry about each outcome and not whether it is the same information, and the measure of Bertschinger and colleagues, BROJA, defines unique information as the minimum of a conditional mutual information over the joint distributions with the same pairwise marginals $P(Y, C)$ and $P(Y, R)$ (Williams and Beer 2010; Bertschinger et al. 2014). We use BROJA because redundancy should mean the same information. SURD, which decomposes the causal influence of lagged variables into the same three kinds of component, needs time structure that the one-shot experiments below do not have (Martínez-Sánchez et al. 2024). Multivariate decompositions remain unsettled, and the analysis stays at two sources.

On the standard cases the implementation returns the known values (Table 2): $Y = C = R$ is redundant by 1 bit, $Y = (C, R)$ with independent bits is unique to each source by 1 bit, the exclusive-or is synergistic by 1 bit, and the conjunction splits into 0.311 bits of redundancy and 0.5 bits of synergy.

Table 2: BROJA decomposition in bits of standard joint distributions of binary variables.

case	redundant	unique to C	unique to R	synergistic
$Y = C = R$	1.000	0.000	0.000	0.000
$Y = C$	0.000	1.000	0.000	0.000
$Y = (C, R)$	0.000	1.000	1.000	0.000
$Y = C \text{ xor } R$	0.000	0.000	0.000	1.000
$Y = C \text{ and } R$	0.311	0.000	0.000	0.500

Plug-in estimates are biased upward. When C , R and Y are independent, the first-order bias of the combination of mutual informations that the synergy contains is $(|C| - 1)(|R| - 1)(|Y| - 1)/(2n \ln 2)$ bits (Treves and Panzeri 1995; Paninski 2003). In simulation the plug-in BROJA synergy of independent variables with alphabets of 3, 3 and 2 is 0.081 bits at 50 responses, 0.043 at 100, 0.011 at 300, 0.0033 at 1,000 and 0.0012 at 3,000, about one and a half times the first-order formula at 100 responses; with binary variables it is 0.020, 0.0086, 0.0046, 0.0010 and 0.0004 bits. The formula puts the sample size at which the bias falls to 0.01 bits at 72 responses for alphabets of 2, 2 and 2, 577 for 3, 2 and 5, 4,617 for 5, 5 and 5 and 15,581 for 7, 7 and 7, to the nearest integer (Figure 4). The noise matters more than the bias in the useful range: for an exclusive-or with 10% flip noise, whose true synergy is 0.531 bits, the standard deviation of the plug-in estimate is 0.154 bits at 50 responses, 0.048 at 300 and 0.018 at 3,000. With noise of that size synergy should not be the only primary endpoint.

6. Designs and what they resolve

6.1 The controlled factorial experiment

Decompose the reaction into modules, source selection and compression Q , interpretive commentary I , affective and social reaction A and editorial operations E , and randomize every viewer-item over the 16 combinations. The cell means give the dividends and the Shapley shares of all four modules by arithmetic. Their sampling variance is known in closed form. A dividend of $|T|$ modules is a signed sum of $2^{|T|}$ cell means, so with N observations spread over 16 cells its variance is $16 \sigma^2 2^{|T|} / N$; the Shapley share of one module has squared coefficient norm $1/3$ and variance $16 \sigma^2 / (3N)$.

A standard error of 0.05 standard deviations of the outcome therefore needs 2,133 observations for a Shapley share, 12,800 for a main-effect dividend, 25,600 for a pair, 51,200 for a triple and 102,400

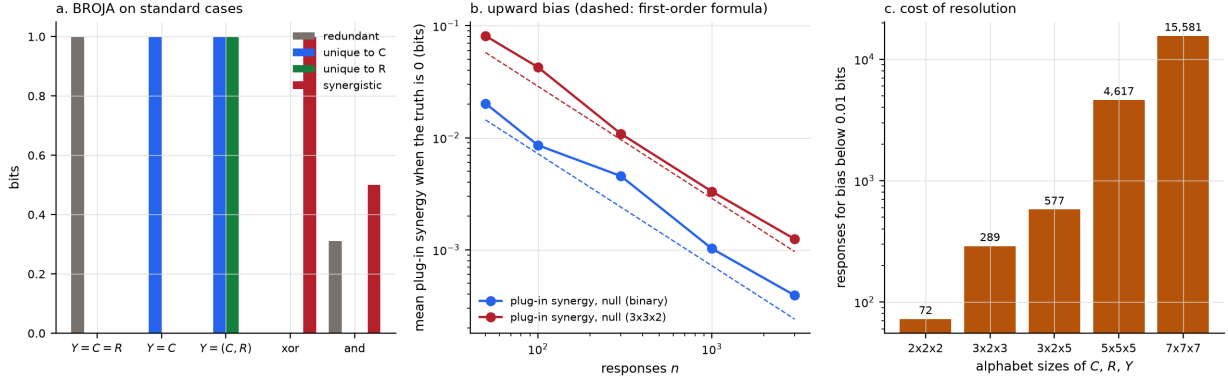


Figure 4: Partial information decomposition. (a) BROJA components of standard cases. (b) Mean plug-in synergy against the number of responses when the truth is zero, for binary variables and for alphabets of 3, 3 and 2; dashed lines are the first-order formula. (c) Responses needed for a bias of 0.01 bits against the alphabet sizes of C , R and Y .

for the four-way dividend. A power of 0.8 against a dividend of 0.2 standard deviations needs 1,047 observations for a Shapley share, 6,279 for a main effect, 12,558 for a pair, 25,116 for a triple and 50,233 for the four-way (Table 3). At 3,000 observations the standard errors are 0.042, 0.103, 0.146, 0.207 and 0.292 standard deviations for the five estimands, at 15,000 they are 0.019, 0.046, 0.065, 0.092 and 0.131, and at 25,000 they are 0.015, 0.036, 0.051, 0.072 and 0.101. A Monte Carlo check with 20,000 replicates of the cell means agrees with the formulas to within 1%. The four-way dividend enters the Shapley share of each module with weight $1/4$, so a design that cannot estimate it leaves a bias of up to a quarter of its size. A half-fraction of eight cells identifies at most eight of the sixteen dividends, so the Shapley share can then be computed only under assumptions about the others (Collins et al. 2009).

Table 3: Observations needed in the 16-cell design, with cells randomized per observation.

estimand	squared coefficient	observations for SE 0.05 SD	observations for 80% power, 0.2 SD
Shapley share	1/3	2,133	1,047
main-effect dividend	2	12,800	6,279
pair dividend	4	25,600	12,558
triple dividend	8	51,200	25,116
four-way dividend	16	102,400	50,233

Viewers and stimuli are sampled, so the outcomes have crossed random effects (Gelman and Hill 2007), and conventional power calculations that ignore them mislead (Judd et al. 2012; Westfall et al. 2014). In a simulation with 600 viewers each seeing five items on 40 stimuli, with variance shares of 0.30 for viewers, 0.10 for stimuli and 0.60 for the residual, the cell-mean estimator of the

Shapley share of I has a standard deviation of 0.0409 against 0.0422 from the formula, so viewer and stimulus effects add sampling variance equal to additive noise. A least-squares estimator that absorbs viewer and stimulus effects has 0.0372, a variance ratio of only 1.21: five items per viewer in randomized cells give little within-viewer information. The sample size should come from a simulation with the planned structure, and the formulas above set its floor.

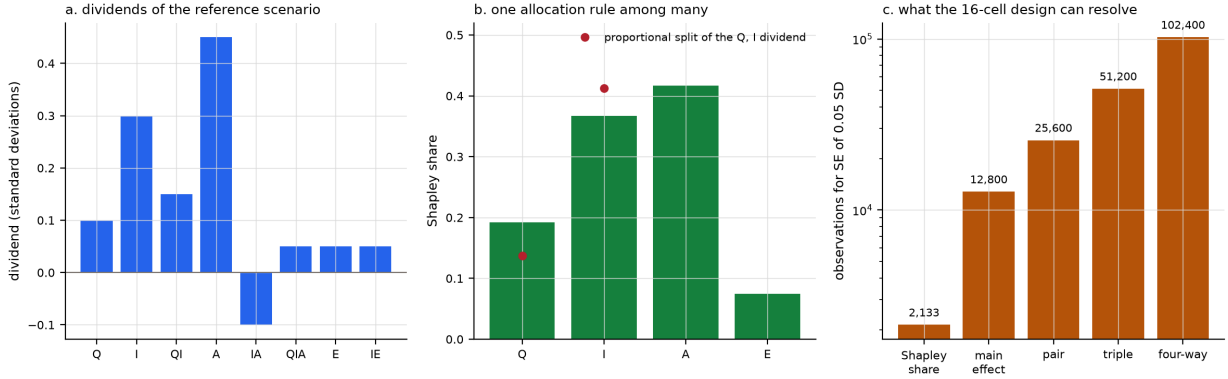


Figure 5: The 16-cell design. (a) Dividends of the reference scenario. (b) Shapley shares and the allocation of the Q , I dividend under a proportional rule. (c) Observations needed for a standard error of 0.05 standard deviations.

6.2 The field experiment

In the field experiment source creators are recruited, eligible videos are randomized to receive a reaction or not, and the originals are measured. Randomization identifies a causal effect of the reaction on the source only for the outcome and the unit at which the assignment varies. With within-creator randomization the unit is the video, and the standard error of the effect on log views with a baseline covariate is $\sigma_w \sqrt{(1 - \rho^2) / (0.25 n (1 - 1/m))}$ for n videos, a within-creator standard deviation σ_w , baseline correlation ρ and m eligible videos per creator. With $m = 10$, $n = 300$, $\sigma_w = 1.0$ and $\rho = 0.7$ the minimum detectable effect at a power of 0.8 and a two-sided level of 0.05 is 0.244 log points, 27.6% in views; a simulation reproduces the standard error (0.0842 against 0.0869). Across the grid with $m = 10$, σ_w from 0.8 to 1.2 and ρ from 0.5 to 0.8, the minimum detectable effect is 33% to 85% at 100 videos, 18% to 43% at 300 and 9.4% to 21% at 1,000; detecting a 10% effect needs between 885 and 4,147 videos. A field experiment of 100 to 300 videos is informative only if the effect on views is large.

Within-creator randomization does not identify the effect on a creator's whole catalogue. Spillover across a creator's catalogue breaks the assumption of no interference (Hudgens and Halloran 2008). Let a reaction to one video raise the other videos of the same creator by δ per unit share of those videos treated, and let the effect on the treated video be τ . Then the within-creator estimate converges to $\tau - \delta / (m - 1)$, which is neither τ nor the total effect $\tau + \delta$ of a reaction to the whole catalogue. With $\tau = 0.10$, $\delta = 0.15$ and $m = 10$ the within-creator estimate is 0.084 in simulation (0.083 analytic) against a total of 0.25. Randomizing at the creator level identifies the total effect, 0.250 in simulation, but with 30 creators the standard deviation of the estimate is 0.142 log points

and the minimum detectable effect is 0.397 log points, 48.7% (Figure 6). Effects on subscribers, catalogue viewing and revenue are creator-level outcomes and share that limit. Staggered reaction dates within the experiment call for event-study estimators that allow heterogeneous effects (Sun and Abraham 2021; Callaway and Sant’Anna 2021). The causal framework is the potential-outcomes one (Rubin 1974; Pearl 2009; Imbens and Rubin 2015).

6.3 Dose

Randomizing the share q of the source that the reaction reproduces over four arms, $q \in \{0.1, 0.3, 0.6, 0.9\}$, lets the combined curve $V_{viewer}(q) + M_{source}(q)$ be estimated at four points; its maximum is located only to the grid unless a parametric curve is assumed. In the simulation outcomes have unit standard deviation per video, the arm $q = 0.3$ is best, the arm $q = 0.6$ is second best and the other two arms are 0.2 and 0.4 standard deviations below the second best. With the best arm 0.1 standard deviations above the second best, the probability of identifying it is 0.71 at 75 videos per arm, 0.81 at 150, 0.89 at 300, 0.95 at 600 and 0.99 at 1,200; with a gap of 0.2 it is 0.89, 0.96 and 0.99 at 75, 150 and 300; with a gap of 0.3 it is 0.96 at 75 and 0.99 at 150. The result is an empirical substitution frontier and not a copyright threshold.

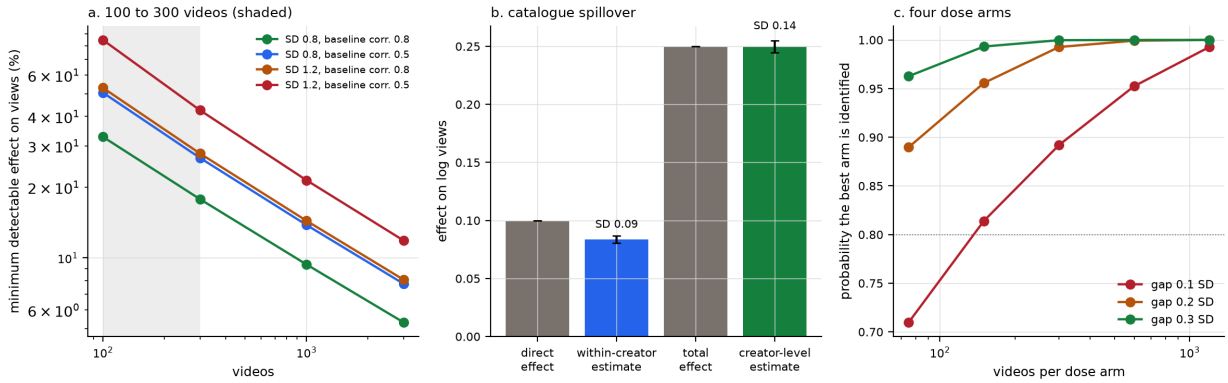


Figure 6: The field experiment. (a) Minimum detectable effect on views against the number of videos for four combinations of within-creator standard deviation and baseline correlation; the shaded band is 100 to 300 videos. (b) Direct effect, within-creator estimate, total effect and creator-level estimate under catalogue spillover, with 30 creators and 10 videos each; bars show twice the standard error of the simulation mean and labels give the standard deviation of one estimate. (c) Probability of identifying the best of four dose arms against videos per dose arm.

6.4 The observational graph

A reaction-source graph, with an edge from reaction R_i to source C_j , can be built from the explicit source links in descriptions, title mentions, creator names, timing and semantic similarity, with the explicit links as a gold standard and a classifier trained on them. It describes and replicates; it does not identify the externality, because reactors choose their sources, and a source that is trending is more likely to be reacted to and to gain views. The YouNiverse dataset comprises metadata from over 136k channels and 72.9M videos published between May 2005 and October 2019, weekly

channel time series and a table of which videos 449M anonymous users commented on (Horta Ribeiro and West 2021), and the YouTube Researcher Program gives eligible researchers scaled access to the public data corpus (YouTube 2026). The pairing step needs a validation sample: a Wilson interval for a precision of 0.95 with a half-width of 0.02 needs 469 pairs, for 0.90 it needs 868 and for 0.98 it needs 223 (Brown et al. 2001). An automated labeler should not decide pairings unchecked.

6.5 Hypotheses for preregistration

Preregistration fixes hypotheses before data (Nosek et al. 2018). The needle model implies that the epistemic gain of commentary that compresses the source rises with the difficulty of the source, $\partial \Delta E_R / \partial D_C \geq 0$, strictly until viewers abandon the source; the model does not cover hedonic gain, which for affective reaction we expect to be unrelated to difficulty. A restream that repeats the source should be a garbling, d_B near zero by the bootstrap test at the nominal 0.025, and an expert reaction should show $Unq_R + Syn > Red$. A reaction that raises reported confidence without raising resolution should show a lower gain through the last term of Proposition 3.

7. Contextual attribution

Let the context space be the product of genre, expertise, audience and platform. A cover is a family of contexts $\{U\}$; each carries a local attribution vector s_U , and overlapping contexts must agree on their overlap through restriction maps $\rho_{U,U \cap V}$. The inconsistency energy is

$$\mathcal{E}(s) = \sum_{U \sim V} \|\rho_{U,U \cap V} s_U - \rho_{V,U \cap V} s_V\|^2,$$

the quadratic form of a cellular-sheaf Laplacian (Hansen and Ghrist 2019). A global attribution summary is justified when $\mathcal{E}$ is zero.

Proposition 5. With identity restriction maps, $\mathcal{E}(s) = s^\top (L \otimes I_d) s$ where L is the Laplacian of the overlap graph. It is zero exactly when s is constant on each connected component. If all contexts are disjoint there are no overlaps and every family of local attributions glues. If the overlap graph is complete on K contexts, $\mathcal{E}(s) = K \sum_U \|s_U - \bar{s}\|^2$, the statistic of the homogeneity test. If the cover consists of the genre strata and the expertise strata of a balanced grid, with each genre stratum and expertise stratum overlapping in one cell, $\mathcal{E}(s) = n_E SS_{\text{genre}} + n_G SS_{\text{expertise}}$, where n_G and n_E are the numbers of genres and expertise levels and SS are the sums of squares of the stratum means about the grand mean, and it vanishes for every pure interaction whose marginals vanish.

Proof. The first statement is the identity $\sum_{U \sim V} (s_U - s_V)^2 = s^\top L s$ and the null space of a graph Laplacian. The disjoint case has an empty edge set. For the complete graph $L = KI - \mathbf{1}\mathbf{1}^\top$. For the marginal cover, $\sum_{g,e} (r_g - c_e)^2$ with r_g the mean of genre g and c_e the mean of expertise level e expands to $n_E \sum_g (r_g - \bar{s})^2 + n_G \sum_e (c_e - \bar{s})^2$ because $\bar{r} = \bar{c} = \bar{s}$, and a pure interaction has $r_g = c_e = \bar{s}$. $\square$

The number of connected components of the overlap graph is the number of separate attribution summaries the cover can justify. The energy is a heterogeneity statistic whose directions of sensitivity are those of the cover. With identity restrictions nothing beyond the graph Laplacian is needed; the sheaf structure matters when restriction maps are not identities, for instance when outcomes on two platforms are measured in different units and must be converted before they are compared.

In a grid of four genres and three expertise levels, three covers were compared on four alternatives of equal total variation, with centered cell effects whose sum of squares is 25 for cell estimates of unit variance: a genre effect, an expertise gradient, a checkerboard interaction with zero marginals and a single deviant cell (Figure 7). The complete cover has one component, the marginal cover has one and the ladder cover, in which cells of one genre at adjacent expertise levels overlap, has four. The complete cover is the homogeneity test and has power 0.95 against all four. The marginal cover has power 0.98 against the genre effect, 0.99 against the gradient, 0.75 against the single cell and 0.05 against the checkerboard: it cannot see a pure interaction. The ladder cover has power 0.05 against the genre effect, which lies in the null space of its Laplacian, 0.70 against the gradient, 0.69 against the single cell and 0.99 against the checkerboard. A report of low inconsistency is therefore a statement about a cover, and the cover has to be reported with the alternatives it can reject.

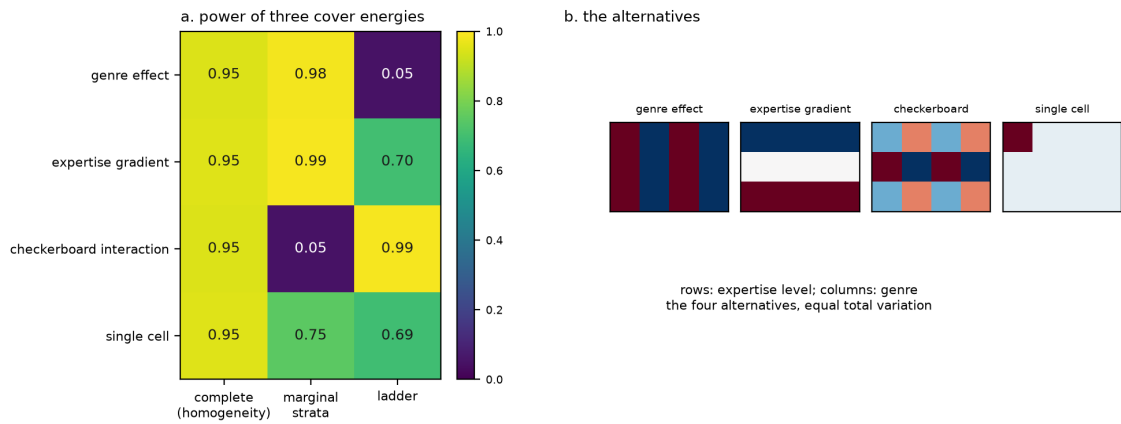


Figure 7: Consistency energy of three covers over a grid of four genres and three expertise levels. (a) Power of the three energies against four alternatives of equal total variation. (b) The alternatives, with columns for genre and rows for expertise level.

8. The attribution signature

The result of the program is a vector and not a share: the viewer-side value components ΔE , ΔH and ΔS , an attention efficiency η_A , the unique and synergistic information Unq_R and Syn , the interaction Γ_{CR} , the module dividends or their Shapley allocation, the source externality ΔM , and the context inconsistency. A statement such as “the reaction supplies a fifth of the value” collapses

the four quantities of section 2 and chooses a weight and an allocation rule that no experiment supplies. Table 4 lists what each component needs.

Table 4: Components of the attribution signature, the design that identifies each, the estimator and its cost.

component	design	estimator	resolution
$\Delta E, \Delta H, \Delta S$	randomized exposure, pre and post forecasts	difference of means; Brier gain with the miscalibration term	1,600 observations for a standard error of 0.05 SD in a two-arm contrast
Γ_{CR} , module dividends	16-cell factorial	Möbius transform of cell means	12,800 to 102,400 observations for 0.05 SD, by order
Shapley allocation	16-cell factorial	Shapley value of the measured game	2,133 observations for 0.05 SD; not identified beyond the dividends
Unq_R, Syn	outcomes with source and reactor content as discrete variables	BROJA with bias and noise reported	577 responses for a bias of 0.01 bits at alphabets 3, 2, 5
garbling	response channels per state and condition	calibrated bootstrap on d_B	1,000 responses per state and condition for power 0.92 at accuracy 0.6
ΔM	randomized reaction on source videos	within-creator or creator-level contrast	28% detectable at 300 videos; 49% at 30 creators
context inconsistency	cover over contexts	consistency energy of the reported cover	set by the cover

9. Limits and generalization

The decision-theoretic results hold for a Bayesian viewer with a fixed utility and the stated cost structure. Proposition 1 does not describe viewers who misread the source, and Proposition 3 treats stated probabilities as reports of beliefs. The needle model has one informative segment, a uniform position and a linear price; its closed form is the point of the example and not an estimate of real viewing. The partial information decomposition is one of several, the multivariate case is unsettled, and its application needs discrete content variables that real reactions rarely supply. The design calculations assume independent outcomes with a stated variance, additive viewer and stimulus effects and, in the field, log views with a given within-creator spread and baseline correlation; the true values are unknown, and the formulas let a pilot replace them. The simulations

use illustrative scenarios and give no estimate of any real reaction, reactor or platform. No data were collected. The quantities are empirical and do not stand in for a copyright test.

The structure is not particular to reaction videos. A news commentary on an article, an explanation of a research paper, a criticism of a film, a remix of a song, a fork of software and a summary of a human work by a machine have the same four quantities, and the designs above apply wherever the source and the derivative can both be observed and exposure can be randomized. Where exposure cannot be randomized, the observational graph of section 6.4 describes the pairs and identifies none of the four quantities without further assumptions.

References

- Bertschinger, Nils, Johannes Rauh, Eckehard Olbrich, Jürgen Jost, and Nihat Ay. 2014. "Quantifying Unique Information." *Entropy* 16 (4): 2161–83. <https://doi.org/10.3390/e16042161>.
- Blackwell, David. 1953. "Equivalent Comparisons of Experiments." *The Annals of Mathematical Statistics* 24 (2): 265–72. <https://doi.org/10.1214/aoms/1177729032>.
- Brier, Glenn W. 1950. "Verification of Forecasts Expressed in Terms of Probability." *Monthly Weather Review* 78 (1): 1–3. [https://doi.org/10.1175/1520-0493\(1950\)078%3C0001:vofeit%3E2.0.co;2](https://doi.org/10.1175/1520-0493(1950)078%3C0001:vofeit%3E2.0.co;2).
- Brown, Lawrence D., T. Tony Cai, and Anirban DasGupta. 2001. "Interval Estimation for a Binomial Proportion." *Statistical Science* 16 (2). <https://doi.org/10.1214/ss/1009213286>.
- Callaway, Brantly, and Pedro H.C. Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics* 225 (2): 200–230. <https://doi.org/10.1016/j.jeconom.2020.12.001>.
- Caplin, Andrew, and Mark Dean. 2015. "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review* 105 (7): 2183–203. <https://doi.org/10.1257/aer.20140117>.
- Caplin, Andrew, Mark Dean, and John Leahy. 2022. "Rationally Inattentive Behavior: Characterizing and Generalizing Shannon Entropy." *Journal of Political Economy* 130 (6): 1676–715. <https://doi.org/10.1086/719276>.
- Collins, Linda M., John J. Dziak, and Runze Li. 2009. "Design of Experiments with Multiple Independent Variables: A Resource Management Perspective on Complete and Reduced Factorial Designs." *Psychological Methods* 14 (3): 202–24. <https://doi.org/10.1037/a0015826>.
- Cover, Thomas M., and Joy A. Thomas. 2006. *Elements of Information Theory*. 2nd ed. Wiley. <https://doi.org/10.1002/047174882x>.
- DeGroot, Morris H., and Stephen E. Fienberg. 1983. "The Comparison and Evaluation of Forecasters." *The Statistician* 32 (1/2): 12. <https://doi.org/10.2307/2987588>.

- Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data Via the EM Algorithm." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 39 (1): 1–22. <https://doi.org/10.1111/j.2517-6161.1977.tb01600.x>.
- Efron, Bradley, and R.J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman and Hall. <https://doi.org/10.1201/9780429246593>.
- Fisher, Alec. 2021. "Unspoken Criticism: Audiovisual Forms of Critique as Fair Use." *The Columbia Journal of Law & the Arts* 44 (2). <https://doi.org/10.52214/jla.v44i2.7825>.
- Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press. <https://doi.org/10.1017/cbo9780511790942>.
- Ghosh, Anirban, and Sarvesh Dutt Tripathi. 2025. "Understanding Viewer Fascination with Reaction Videos: A Discourse Analysis of YouTube Comments." *Media Watch* 16 (2): 230–49. <https://doi.org/10.1177/09760911241313236>.
- Gneiting, Tilmann, and Adrian E Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102 (477): 359–78. <https://doi.org/10.1198/016214506000001437>.
- Grabisch, Michel. 2016. *Set Functions, Games and Capacities in Decision Making*. Springer. <https://doi.org/10.1007/978-3-319-30690-2>.
- Hansen, Jakob, and Robert Ghrist. 2019. "Toward a Spectral Theory of Cellular Sheaves." *Journal of Applied and Computational Topology* 3 (4): 315–58. <https://doi.org/10.1007/s41468-019-00038-7>.
- Harsanyi, John C. 1959. "A Bargaining Model for the Cooperative n-Person Game." In *Contributions to the Theory of Games, Volume IV*. Princeton University Press. <https://doi.org/10.1515/9781400882168-019>.
- Heskes, Tom, Evi Sijben, Ioan Gabriel Bucur, and Tom Claassen. 2020. "Causal Shapley Values: Exploiting Causal Knowledge to Explain Individual Predictions of Complex Models." In *arXiv Preprint, arXiv*. <https://doi.org/10.48550/ARXIV.2011.01625>.
- Horta Ribeiro, Manoel, and Robert West. 2021. "YouNiverse: Large-Scale Channel and Video Metadata from English-Speaking YouTube." *Proceedings of the International AAAI Conference on Web and Social Media* 15: 1016–24. <https://doi.org/10.1609/icwsm.v15i1.18125>.
- Hudgens, Michael G, and M. Elizabeth Halloran. 2008. "Toward Causal Inference With Interference." *Journal of the American Statistical Association* 103 (482): 832–42. <https://doi.org/10.1198/016214508000000292>.
- Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press. <https://doi.org/10.1017/cbo9781139025751>.

- Janzing, Dominik, Lenon Minorics, and Patrick Blöbaum. 2019. "Feature Relevance Quantification in Explainable AI: A Causal Problem." In *arXiv*. Preprint, arXiv. <https://doi.org/10.48550/ARXIV.1910.13413>.
- Judd, Charles M., Jacob Westfall, and David A. Kenny. 2012. "Treating Stimuli as a Random Factor in Social Psychology: A New and Comprehensive Solution to a Pervasive but Largely Ignored Problem." *Journal of Personality and Social Psychology* 103 (1): 54–69. <https://doi.org/10.1037/a0028347>.
- Lundberg, Scott, and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions." In *arXiv*. Preprint, arXiv. <https://doi.org/10.48550/ARXIV.1705.07874>.
- Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. "Rational Inattention: A Review." *Journal of Economic Literature* 61 (1): 226–73. <https://doi.org/10.1257/jel.20211524>.
- Martínez-Sánchez, Álvaro, Gonzalo Arranz, and Adrián Lozano-Durán. 2024. "Decomposing Causality into Its Synergistic, Unique, and Redundant Components." *Nature Communications* 15 (1). <https://doi.org/10.1038/s41467-024-53373-4>.
- Matějka, Filip, and Alisdair McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review* 105 (1): 272–98. <https://doi.org/10.1257/aer.20130047>.
- Murphy, Allan H. 1973. "A New Vector Partition of the Probability Score." *Journal of Applied Meteorology* 12 (4): 595–600. [https://doi.org/10.1175/1520-0450\(1973\)012%3C0595:anvpot%3E2.0.co;2](https://doi.org/10.1175/1520-0450(1973)012%3C0595:anvpot%3E2.0.co;2).
- Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences* 115 (11): 2600–2606. <https://doi.org/10.1073/pnas.1708274114>.
- Paninski, Liam. 2003. "Estimation of Entropy and Mutual Information." *Neural Computation* 15 (6): 1191–253. <https://doi.org/10.1162/089976603321780272>.
- Pearl, Judea. 2009. *Causality*. 2nd ed. Cambridge University Press. <https://doi.org/10.1017/cbo9780511803161>.
- Rota, Gian-Carlo. 1964. "On the Foundations of Combinatorial Theory I. Theory of Möbius Functions." *Zeitschrift Für Wahrscheinlichkeitstheorie Und Verwandte Gebiete* 2 (4): 340–68. <https://doi.org/10.1007/bf00531932>.
- Rubin, Donald B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66 (5): 688–701. <https://doi.org/10.1037/h0037350>.
- Self, Steven G., and Kung-Yee Liang. 1987. "Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions." *Journal of the American Statistical Association* 82 (398): 605–10. <https://doi.org/10.1080/01621459.1987.10478472>.

- Shapley, L. S. 1953. "A Value for n-Person Games." In *Contributions to the Theory of Games, Volume II*. Princeton University Press. <https://doi.org/10.1515/9781400881970-018>.
- Sims, Christopher A. 2003. "Implications of Rational Inattention." *Journal of Monetary Economics* 50 (3): 665–90. [https://doi.org/10.1016/s0304-3932\(03\)00029-1](https://doi.org/10.1016/s0304-3932(03)00029-1).
- Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics* 225 (2): 175–99. <https://doi.org/10.1016/j.jeconom.2020.09.006>.
- Treves, Alessandro, and Stefano Panzeri. 1995. "The Upward Bias in Measures of Information Derived from Limited Data Samples." *Neural Computation* 7 (2): 399–407. <https://doi.org/10.1162/neco.1995.7.2.399>.
- Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453–58. <https://doi.org/10.1126/science.7455683>.
- Westfall, Jacob, David A. Kenny, and Charles M. Judd. 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): 2020–45. <https://doi.org/10.1037/xge0000014>.
- Williams, Paul L., and Randall D. Beer. 2010. "Nonnegative Decomposition of Multivariate Information." In *arXiv*. Preprint, arXiv. <https://doi.org/10.48550/ARXIV.1004.2515>.
- Witter, R. Teal, Álvaro Parafita, Tomas Garriga, et al. 2026. "Exactly Computing Do-Shapley Values." In *arXiv*. Preprint, arXiv. <https://doi.org/10.48550/ARXIV.2602.07203>.
- YouTube. 2026. "YouTube Researcher Program." <https://research.youtube/>.