TWOCHANNEL: SEPARATING MISSING FROM CORRUPTED MODALITIES IN MULTIMODAL FUSION

Anonymous authors
Paper under double-blind review

ABSTRACT

A multimodal system in deployment decides from whatever channels arrived, and a channel that never arrived and one that arrived corrupted call for opposite repairs, a new acquisition or a recalibration. Existing fusion rules write both failures onto one non-negative scalar per channel, whose margin narrows as the damage deepens, and neither filling nor a trust fitted to task loss recovers the distinction. We propose TwoChannel, a fusion rule that carries absence and unreliability on a support mass and a conflict mass rather than on one weight per channel. Prototype decomposition reads the sub-views of each channel onto a shared dictionary and splits what the channel asserts from where its own parts disagree, so each channel enters fusion as an assertion and a contradiction. Provenance-typed transport fills the slots an absent channel leaves along compatibility edges and writes restored evidence into the support mass only, so a restoration can raise support but cannot manufacture conflict. A per-prototype trust product combines a channel's internal dispersion with its agreement with the other present channels and is read from the instance alone, with no label naming the corrupted channel. We prove that neither an absent channel nor restored evidence adds a conflict term, so the separation holds by construction rather than by estimation. On Syn-MOSEI, Syn-IEMOCAP and Syn-ADNI, a probe separating a dropped from a damaged channel reaches 0.9960 off this readout against a raw-input ceiling of 1.0000, where mean filling leaves it at 0.4995, near the 0.5019 of a random-feature control. On the primary cohort it keeps that distinction at no accuracy cost, matching plain fusion's 0.9383 at complete input where the scalar gate reads 0.7883.

1 INTRODUCTION

A multimodal system in deployment decides from whatever channels arrived, and channels arrive in two broken states: some are missing outright, and some are present but corrupted (Li et al., 2025; Fu et al., 2025; Luo et al., 2025; Zeng et al., 2022; Zuo et al., 2023; Huang et al., 2022). The two failures call for opposite responses: a channel that never arrived is an acquisition failure and is repaired upstream, while a channel that arrived wrong is a calibration failure and is repaired at the sensor. Which of the two occurred is therefore not a detail of the input but the content of the judgment a downstream policy has to make. Everything that policy can read is the single fused vector the fusion rule hands it. A rule that cannot keep the two apart inside that vector routes an operator to the wrong repair with no sign of the mistake (Figure 1).

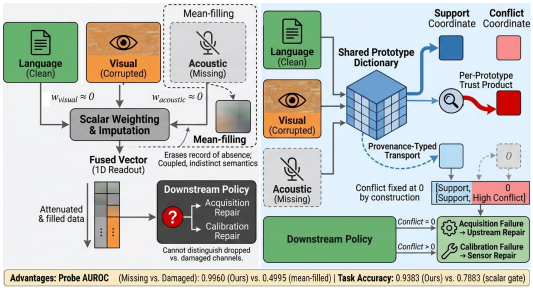

Figure 1: **Left:** one attenuation scalar and a mean filling hand the policy a vector in which a dropped and a damaged channel look alike. **Right:** TwoChannel keeps support and conflict apart; a missing channel's conflict is zero.

One line of work reads each channel's usefulness as a single non-negative scalar and attenuates the channel by it, learning that weight end to end from the task (Wu et al., 2025; Meng et al., 2026; Zhang et al., 2023). This keeps the fusion differentiable and lets one model serve every degradation pattern without a separate branch per pattern. A second line reconstructs the evidence an absent channel would have supplied, learning tag-conditioned and hierarchical recovery that fills the empty slot before fusion (Zeng et al., 2022; Li et al., 2024), and the fusion sees a complete instance. A

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

third line estimates how far each channel should be trusted, reading an evidential belief mass or a calibrated reliability score from the instance itself (Huang et al., 2022; Han et al., 2021). A parallel strand represents each channel through learned prototypes and propagates evidence over a compatibility graph, supplying the machinery that names assertions from different channels on common coordinates (Liu et al., 2025; Li et al., 2023b). Together these lines raise accuracy across missing-modality benchmarks and supply the mechanisms this work assembles. Despite these advances, all four deliver their result through a readout of the same shape.

The four lines differ in what they write into that readout and agree on its shape: one non-negative coordinate per channel. On that coordinate a channel that never arrived and a channel that arrived corrupted are both driven toward zero, and what separates them is only the distance the damage has not yet closed. Restoration compounds this rather than relieving it, since filled content re-enters through the coordinates that held observed content. Nor can a trust weight fitted against task loss say which failure occurred, since nothing at serving time labels the degraded channel. What reaches the downstream policy is then one number per channel that cannot say whether to request a new acquisition or a recalibration. Absence and corruption are therefore coupled onto a single attenuation axis that no gain in accuracy can split, and that each of the four lines above inherits.

To address these coupled failures, we propose TwoChannel, a fusion rule that carries a support coordinate and a conflict coordinate through fusion in place of one attenuation scalar. The key insight is that an absent channel's contribution to the conflict mass is fixed at exactly zero by construction rather than by estimation. TwoChannel has three parts: a prototype decomposition that reads each channel onto a shared dictionary as a support mass and a conflict mass; a provenance-typed transport along compatibility edges that fills the slots an absent channel leaves and is confined to the support mass; and a per-prototype trust product read from a channel's own dispersion and its agreement with the other present channels, with no degradation label anywhere on the path. One objective couples the three parts, and the readout carrying both masses is delivered together with the presence record, so the decision can still read which of the two failures occurred. Our contributions are as follows:

- We identify that the incumbent fusion families, whatever they write into the fused vector, write it onto one non-negative coordinate per channel, on which the margin between an absent channel and a corrupted one narrows as the damage deepens.
- We propose TwoChannel, which carries absence on a support coordinate and unreliability on a conflict coordinate, restores an absent channel by transport confined to the support mass, and resolves per-prototype trust from the instance alone.
- A probe asked to tell a dropped channel from a damaged one reaches 0.9960 off our readout against 0.4995 for a mean-filled representation and a raw-input ceiling of 1.0000, while task accuracy reads 0.9383 against the scalar gate's 0.7883 at complete input.

2 MOTIVATION

To understand what a fused readout still says about a failed channel, we measure what each fusion family hands downstream, revealing three challenges.

Challenge #1: One attenuation axis. Gated fusion summarises a channel's usefulness as one non-negative weight and attenuates the fusion by it. Figure 2 (left) measures how far a dropped channel sits from a damaged one in what each one-coordinate family hands downstream, in units of how far the damage alone moved the same instance, over 12 damage levels. Every family falls across the sweep and the gate lies below all of them, from 1.3795 at the lightest damage to 0.0001 at the heaviest, against a raw input that reads 6.4459 down to 1.4557. The root cause is structural: an absent channel drives the weight to zero and a damaged one drives it down, so the two states share one coordinate by construction. *How can a fused readout keep absence and corruption on coordinates a decision can still read apart?*

Challenge #2: An erased absence record. The natural recourse is to fill the absent channel in. Figure 2 (middle) asks a probe which channel was never there, reading only what is handed on, with and without a learned filling, on two absence processes that are not interchangeable and are never averaged. On the primary cohort (Syn-MOSEI, Section 5), without filling the probe stays at 0.9949 and above; after filling it falls to between 0.5227 and 0.7756 once channels are withheld independently, while a fixed construction with exactly one channel absent leaves it at 0.9991. The

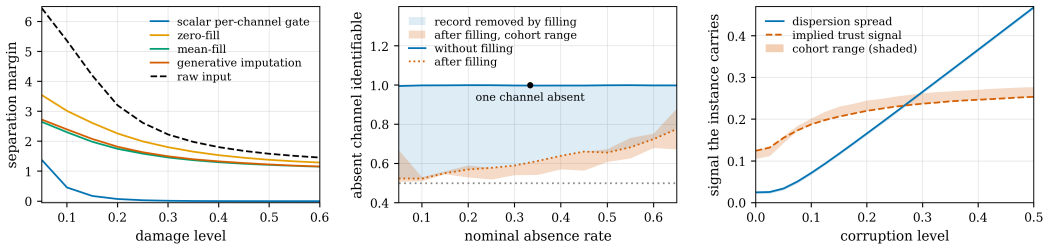

Figure 2: Separation margin against damage (left), absent-channel identifiability with what a filling removes shaded (middle), and the signals an instance carries (right). Details in Appendix A.2.

root cause is structural: a filled entry is delivered through the coordinates that held observed content, and whether the absence stays readable is a property of the filling and of the absence process rather than of the rule that consumes it. *How can an absent channel be restored so that the record of the restoration is what the fusion rule governs?*

Challenge #3: No degradation label. The natural supervision for a reliability weight is a label naming the damaged channel. Figure 2 (right) measures what the instance does carry instead: how far its channels disagree in observable dispersion, and how far the internal-dispersion factor they imply spreads, over 21 corruption levels. Both signals are present and both grow with the damage, the dispersion spread running from 0.0248 to 0.4683 and the implied trust signal from 0.1244 to 0.2537, yet nothing in the instance says which reading is correct. The root cause is structural: reliability is a property of the measurement, and a deployment delivers content and a presence record only, so a weight fit against task loss alone has nothing that resolves it below the channel. *How can a per-prototype reliability be resolved from the content and the presence record alone?*

3 Problem Formulation

To address the three challenges of Section 2, we formulate deployment-time fusion as a constrained risk-minimization problem over a fusion rule β and a decision map h:

$$\min_{\beta \in \mathcal{G},\, h} \; \mathbb{E}\big[\ell\big(h(\Phi_\beta(x, A)),\, y\big)\big], \tag{1}$$

where an instance is a pair (x, A) of channel content x and a presence record A whose indicator a_m marks whether channel m arrived, with $\sum_m a_m \geq 1$; $\Phi_\beta(x, A)$ is the readout, all the decision sees; and ℓ scores h against the target y. The admissible class $\mathcal{G}$ carries the content of the problem. Writing Φ^{abs} and $\Phi^{\mathrm{dam}}(\delta)$ for the readouts of one instance with a given channel absent and with that same channel damaged at level δ, condition (a) requires a margin the damage cannot close,

$$\inf_{\delta} \; \big\|\Phi^{\mathrm{abs}} - \Phi^{\mathrm{dam}}(\delta)\big\| \;\geq\; \varepsilon \;>\; 0. \tag{2}$$

Condition (b) requires that whatever β supplies for an absent channel leave every present channel's contribution unchanged, and that A be delivered with Φ_β. Condition (c) requires β to be a function of x and A alone, with no side annotation of which channel is degraded. Details in Appendix A.3.

Condition (a) answers Challenge #1: a margin that does not shrink with δ is what one attenuation axis gives up. Condition (b) answers Challenge #2: it puts a restoration's record under the rule, not the filling. Condition (c) answers Challenge #3: reliability is read from the instance.

4 Method

4.1 Overall Framework

Figure 3 traces how TwoChannel reads an instance as its content x and presence record A and decomposes every channel onto a dictionary P, producing an evidence graph of channel-to-prototype assertions. Prototype decomposition turns it into a support and a conflict mass per prototype, provenance-typed transport fills the slots an absent channel leaves, writing into the support mass alone, a trust product splits each present channel's mass between the two, and h scores it.

The three are composed rather than chosen because what a deployment wants pulls two ways: a filling buys accuracy and writes the absence record out, while attenuation keeps the record and gives

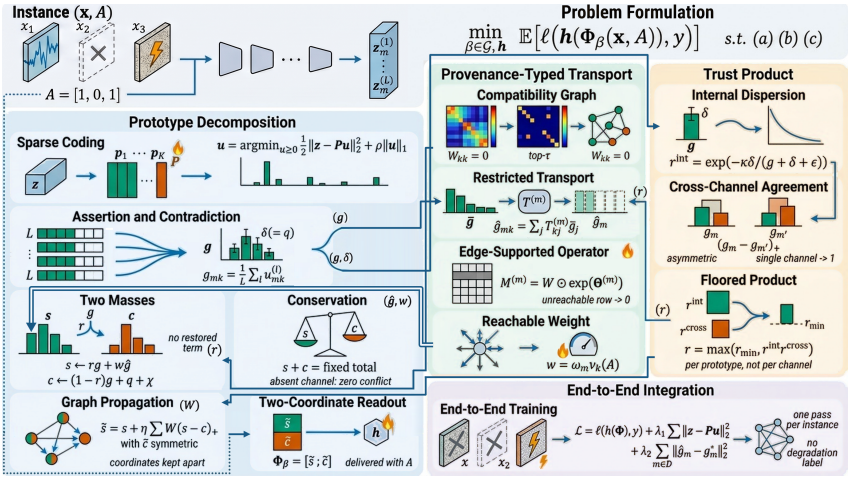

Figure 3: Overall framework of TwoChannel. (**Left**) prototype decomposition of (x, A) into the two masses and the two-coordinate readout; (**Middle**) provenance-typed transport, on the support side alone; (**Right**) the per-prototype trust product; (**Bottom**) end-to-end training.

up accuracy. The composition separates where evidence is written from how strongly it is weighted: a restoration reaches the support mass and never the conflict one, and the trust splits an assertion between them with no label. The readout is a pair of masses, not a weighted sum.

4.2 Prototype Decomposition

The left panel of Figure 3 shows how prototype decomposition answers Challenge #1: it reads each channel onto the dictionary and delivers the two coordinates condition (a) requires, as sub-views are decomposed, assertion is split from contradiction, and the two conserved masses reach the readout.

Sparse Coding. The dictionary P holds K unit-norm columns in $\mathbb{R}^d$ and names assertions from different channels on common coordinates: a column p_k carries the same meaning whichever channel asserts on it. With the encoder of channel m exposed as $L \geq 2$ pooled sub-views $z_m^{(1)}, \ldots, z_m^{(L)}$, one non-negative sparse program is solved per sub-view,

$$u_m^{(l)} = \arg\min_{u \geq 0} \ \tfrac{1}{2}\left\| z_m^{(l)} - Pu \right\|_2^2 + \rho \|u\|_1, \qquad \|p_k\|_2 = 1 \ \ \forall k, \tag{3}$$

where $u_m^{(l)}$ is the code of sub-view l and ρ sets the sparsity penalty. Throughout, ϵ is a fixed positive constant that keeps a denominator away from zero.

Assertion and Contradiction. From the codes we compute a channel's assertion on a prototype as the agreement of its sub-views and its contradiction as their dispersion,

$$g_{mk} = \tfrac{1}{L} \sum_l u_{mk}^{(l)}, \tag{4}$$

where g_{mk} is how much channel m asserts prototype k and the contradiction mass is its dispersion $q_{mk} = \delta_{mk} = \frac{1}{L} \sum_l |u_{mk}^{(l)} - g_{mk}|$: a prototype the channel asserts in some parts while withholding in others is contested, and one no part of it asserts is silent.

Two Masses. Each present channel's assertion is then split by its trust. The trusted part, together with whatever a restoration supplies for an absent channel, forms the support mass,

$$s_k = \sum_m a_m \, r_{mk} \, g_{mk} \ + \ \sum_m (1 - a_m) \, w_{mk} \, \hat{g}_{mk}. \tag{5}$$

The distrusted part, the channel's own contradiction and the disagreement between two present channels form the conflict mass,

$$c_k = \sum_m a_m \left[(1 - r_{mk}) \, g_{mk} + q_{mk} \right] \ + \ \sum_{m < m'} a_m a_{m'} \, \chi_{mm'k}, \tag{6}$$

where $\chi_{mm'k} = \min(g_{mk}, g_{m'k}) \, (1 - \sigma_{mm'})_+$ is the pairwise disagreement of two present channels through the cosine $\sigma_{mm'}$ between their assertion vectors.

Conservation. Because $r_{mk}g_{mk} + (1 - r_{mk})g_{mk} = g_{mk}$ for every present channel, the observed assertion is conserved and merely partitioned, which yields the identity

$$s_k + c_k = \sum_m a_m \big(g_{mk} + q_{mk}\big) + \sum_{m<m'} a_m a_{m'} \chi_{mm'k} + \sum_m (1 - a_m)\, w_{mk}\, \hat{g}_{mk}, \tag{7}$$

where the right side fixes the total the two masses partition within one instance, and every conflict term carries a factor a_m: an absent channel contributes exactly zero conflict.

Graph Propagation. Because prototypes are interdependent, we refine the two masses once over the prototype compatibility graph W of Eq. 10 with the coordinates kept apart,

$$\tilde{s}_k = s_k + \eta \sum_j W_{kj}\, (s_j - c_j)_+, \qquad \tilde{c}_k = c_k + \eta \sum_j W_{kj}\, (c_j - s_j)_+, \tag{8}$$

where η is the propagation strength. A prototype supported by its compatible neighbours gains support, while one whose neighbours are themselves contested gains conflict. The support mass is deliberately not renormalized over present channels, since renormalization would map total assertion back to a constant and restore the one-parameter behaviour of the gated family.

Two-Coordinate Readout. The composed readout is the pair of propagated masses, obtained by concatenating the support and conflict coordinates,

$$\Phi_\beta(x, A) = [\,\tilde{s}\,;\,\tilde{c}\,], \qquad \text{scored by } h(\Phi_\beta), \tag{9}$$

where $\tilde{s}$ and $\tilde{c}$ are the support and conflict masses of Eq. 8 after a single graph propagation, and h is a small head fitted on the readouts. Prototype decomposition therefore hands on the pair Φ_β, which is the readout the decision map h scores and the object the next two components write into.

4.3 Provenance-Typed Transport

The middle panel of Figure 3 shows how provenance-typed transport answers Challenge #2: it fills an absent channel's slots and writes to the support mass alone, as condition (b) asks, with a compatibility graph, transport restricted to reachable slots, and a weight setting its strength.

Compatibility Graph. We estimate compatibility from co-assertion on complete training instances and sparsify it, leaving transport restricted to structure the data exhibits,

$$W_{kj} = \left[\frac{\sum_n \sum_m g^{(n)}_{mk} g^{(n)}_{mj}}{\sqrt{\sum_n \sum_m (g^{(n)}_{mk})^2}\, \sqrt{\sum_n \sum_m (g^{(n)}_{mj})^2}} \right]_\tau, \qquad W_{kk} = 0, \tag{10}$$

where $[\cdot]_\tau$ keeps the largest τ entries of each row and zeroes the rest, and $g^{(n)}_{mk}$ is the assertion on training instance n; τ alone controls the graph's sparsity.

Restricted Transport. For an absent channel the restored assertion is obtained by transport over what the present channels assert, averaged over the present set,

$$\hat{g}_{mk} = \sum_j T^{(m)}_{kj}\, \bar{g}_j, \qquad \bar{g}_j = \frac{\sum_{m'} a_{m'}\, \mathbf{1}[m' \text{ covers } j]\, r_{m'j}\, g_{m'j}}{\max\big(1, \sum_{m'} a_{m'}\big)}, \tag{11}$$

where $\hat{g}_{mk}$ is the restored assertion and $\bar{g}_j$ is the present-channel evidence on prototype j, read on the prototypes each present channel covers.

Edge-Supported Operator. The transport operator is derived on compatibility edges alone, inventing no restored evidence for a prototype unreachable from what is present,

$$T^{(m)}_{kj} = \frac{\mathbf{1}[\bar{g}_j > 0]\, M^{(m)}_{kj}}{\sum_{j'} \mathbf{1}[\bar{g}_{j'} > 0]\, M^{(m)}_{kj'}} \text{ when } \sum_{j'} \mathbf{1}[\bar{g}_{j'} > 0]\, M^{(m)}_{kj'} > 0, \text{ else } 0, \tag{12}$$

where $M^{(m)} = W \odot \exp(\Theta^{(m)})$ carries a per-channel learned modulation $\Theta^{(m)}$ of the graph, and the zero branch sets an unreachable row to zero rather than dividing by zero.

Reachable Weight. The weight of restored evidence is the reachable fraction of a prototype's incoming compatibility mass, scaled by a per-channel gain, which lets restoration enter only where the present channels reach,

$$w_{mk} = \omega_m \cdot \nu_k(A), \qquad \nu_k(A) = \frac{\sum_j W_{kj}\, \mathbf{1}[\bar{g}_j > 0]}{\sum_j W_{kj} + \epsilon}, \tag{13}$$

where ω_m is the learned transport gain and $\nu_k(A)$ is the reachable fraction. A prototype with no reachable support receives $w_{mk} = 0$ and no restored entry. The one-sidedness of restoration is a property of the equations: $\hat{g}_{mk}$ enters the support sum of Eq. 5 under the factor $(1 - a_m)$, Eq. 6 carries no restored term, and none is claimed on a prototype Eq. 13 leaves unreachable. Restoration moves the support mass alone, and the readout is delivered as the pair (Φ_β, A): which channel was absent is read from the record, not inferred from the restored values. Details in Appendix A.4.

4.4 Trust Product

The right panel of Figure 3 shows how the trust product answers Challenge #3: it resolves a per-prototype reliability from the instance alone, as condition (c) asks, by combining an internal-dispersion factor with a cross-channel agreement factor into one floored product per prototype.

Internal Dispersion. The first factor is resolved from how far a channel's own parts disagree about a prototype, measured relative to how much the channel asserts it,

$$r^{\text{int}}_{mk} = \exp\!\left(-\kappa\, \frac{\delta_{mk}}{g_{mk} + \delta_{mk} + \epsilon}\right), \tag{14}$$

where κ is the trust-product temperature. A channel whose assertion on k is internally inconsistent is distrusted on k whether that assertion is large or small.

Cross-Channel Agreement. The second factor is resolved from how far this channel's assertion stands from the other present channels, with the single-channel case given a value of its own,

$$r^{\text{cross}}_{mk} = \left(1 - \frac{\sum_{m' \neq m} a_{m'}\, (g_{mk} - g_{m'k})_+}{\sum_{m' \neq m} a_{m'}\, g_{mk} + \epsilon}\right)_+ \quad \text{if} \sum_{m' \neq m} a_{m'} \geq 1, \quad \text{else } 1, \tag{15}$$

where the second branch gives the value directly when no other present channel exists. The value is not symmetric in m and m': a channel that asserts a prototype the others do not is distrusted on it, and a channel that asserts less than the others is not.

Floored Product. The trust is the floored product of the two factors. With the floor $r_{\min} > 0$ a channel is never fully erased and its assertion is partitioned within Eq. 7 rather than discarded,

$$r_{mk} = \max\!\big(r_{\min},\, r^{\text{int}}_{mk} \cdot r^{\text{cross}}_{mk}\big), \tag{16}$$

where $r_{\min}$ is the trust floor, and the product resolves trust per prototype rather than per channel. A corruption uniform across every part of a channel and also agreeing with every other present channel is invisible to both factors by construction. Details in Appendix A.5.

4.5 End-to-End Integration

End-to-End Training. One objective couples the three components, supervising the restored-evidence path and the low-trust regime in a single loss. Each training channel is removed or corrupted independently at a fixed rate, and we minimize a task risk with a reconstruction term that keeps the dictionary faithful and a transport term supervised by the assertions of removed channels,

$$\mathcal{L} = \mathbb{E}\big[\ell(h(\Phi_\beta), y)\big] + \lambda_1 \sum_m a_m \sum_l \big\| z^{(l)}_m - P u^{(l)}_m \big\|_2^2 + \lambda_2 \sum_{m \in D} \big\| \hat{g}_m - g^*_m \big\|_2^2, \tag{17}$$

where ℓ scores the decision $h(\Phi_\beta)$ against the target y, D is the set of channels removed and g^*_m their withheld assertions, and λ_1, λ_2 weight the two auxiliary terms. The dictionary, the transport operators, the gains and the head are fitted, each by its own rule rather than one joint gradient step, and every other constant is fixed. Details in Appendix A.6.

Table 1: Separation probe ↑, read off each system's fused representation on matched pairs. The raw-input row is the ceiling and the control row the floor. Bold marks this work.

Representation	Syn-MOSEI	Syn-IEMOCAP	Syn-ADNI	Avg
Random-feature control	0.491689	0.491689	0.522188	0.501855
Raw-input probe (ceiling)	1.000000	1.000000	1.000000	1.000000
Gated fusion, one scalar per channel	0.520800	0.530633	0.536406	0.529280
Generative imputation then fusion	0.600522	0.650144	0.470469	0.573712
Zero-fill then fusion	0.699022	0.716556	0.585313	0.666964
Mean-fill then fusion	0.492544	0.523467	0.482500	0.499504
TwoChannel (ours)	**0.998711**	**0.999567**	**0.989844**	**0.996041**

Framework Summary. TwoChannel carries absence and unreliability as two separately readable masses from Eq. 3 through to the readout of Eq. 9. Restoration is admitted to the support mass alone, and the presence record it acted on is delivered with the readout.

5 Experiments

5.1 Implementation Details

Datasets. Readings are taken on three synthetic two-class cohorts with a known prototype code: Syn-MOSEI, Syn-IEMOCAP and Syn-ADNI carry the channel sets of CMU-MOSEI (Bagher Zadeh et al., 2018), IEMOCAP (Busso et al., 2008) and ADNI (Mueller et al., 2005), the last with its normal-control and Alzheimer's-disease groups, and Syn-MOSEI is primary. A channel is made absent by removing it from the presence record and damaged by perturbing it, at three (absence, damage) working points, $(0.1, 0.1)$, $(0.3, 0.2)$ and $(0.5, 0.4)$. Details in Appendix A.7.

Baselines. The baselines are a gated fusion carrying one non-negative scalar per channel, a learned filling of absent channels, zero-filling, mean-filling, a complete-modality reference that never sees the degradation and a random-feature control, all re-implemented and run on the same cohorts under the same degradations; two of them are transductive and this work is not. Details in Appendix A.8.

Evaluation Metrics. The task readings are accuracy ↑, macro F1 ↑ and expected calibration error (ECE) ↓. The mechanism readings are a separation probe ↑ read as an area under the receiver operating characteristic curve (AUROC), the restored and conflict shares of the two masses, a trust rank agreement ↑, and a prototype attribution offset ↓. Details in Appendix A.9.

Hyperparameters. All cohorts and working points share $M = 3$ channels, $L = 4$ sub-views of dimension $d = 32$, $K = 24$ prototypes, $\rho = 0.055$, $\kappa = 3$, $r_{\min} = 0.05$, $\tau = 5$, $\eta = 0.35$, $\lambda_1 = 0.1$, $\lambda_2 = 0.25$ and $\epsilon = 10^{-8}$. Where a reading fixes neither rate, channels are removed with probability 0.2 and corrupted with probability 0.5 at a level of 0.2. Details in Appendix A.10.

5.2 Main Results

Separation at the Readout. Table 1 reports what a decision can still read off each fused representation. The distinction survives this work's readout at 0.9960, while the scalar-gated family reads 0.5293 and no other fused row reaches 0.70 on average, which leaves every other family nearer the 0.5019 a random-feature control reads than the 1.0000 a raw input allows. Among the fillings the one that decides worst keeps the most: zero-filling, the lowest of the three along the absence sweep on every cohort, keeps the distinction at 0.6670, while the learned and mean fillings, which decide better, leave 0.5737 and 0.4995. The filling that restores least leaves the most of the absence record, which is the erased record of Challenge #2 read off the fused representation rather than the input.

Task Accuracy at Complete Input. Table 2 reports what each system decides when nothing is degraded. This work reads 0.9383 against the scalar-gated family's 0.7883 on the primary cohort, a gap of 15.0 points, and stays ahead of that family by 3.3 and 7.5 points on the other two. With nothing absent, the filling families

Table 2: Accuracy ↑ at complete input; the fillings and the reference coincide as plain fusion.

System	Syn-MOSEI	Syn-IEMOCAP	Syn-ADNI
Plain fusion	0.9383	0.9067	0.8438
Gated fusion	0.7883	0.8933	0.7937
TwoChannel (ours)	**0.9383**	**0.9267**	**0.8688**

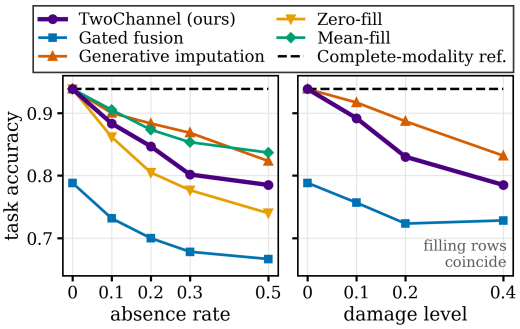

Figure 4: Task accuracy ↑ on the primary cohort, each system fitted where it is read. (**Left**) channels absent; (**Right**) channels damaged.

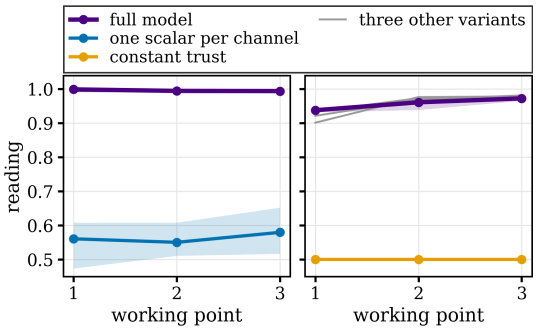

Figure 5: Each variant against the full model at three working points, cohort mean, range shaded. (**Left**) separation probe ↑; (**Right**) trust rank ↑.

Table 3: Each component removed in turn at the middle working point; full model in bold.

Ablated component	Accuracy ↑	Probe ↑	Trust rank ↑
One scalar per channel	0.7761	0.5502	0.9731
TwoChannel (full model)	**0.7615**	**0.9944**	**0.9610**
Imputation for transport	0.7572	0.9961	0.9731
Constant trust	0.7899	0.9969	0.5000
No graph propagation	0.7708	0.9961	0.9763
Residual-only trust	0.7845	0.9928	0.9691

Table 4: Restored share of the support mass and conflict share of the total mass, each at an absence or damage level of 0.4.

	Absence 0.4		Damage 0.4	
Cohort	Restored	Conflict	Restored	Conflict
Syn-MOSEI	0.3651	0.5554	0.0000	0.8671
Syn-IEMOCAP	0.4272	0.4646	0.0000	0.8328
Syn-ADNI	0.4324	0.4937	0.0000	0.8414

and the complete-modality reference reduce to one plain fusion, the row that isolates what the second coordinate costs: this work sits 0.0000, +0.0200 and +0.0250 from it, while the gate gives up between 0.0134 and 0.1500 to it with no channel damaged. The gate pays this price on the decision and keeps no separation for it, 0.5293 on the probe; this work pays none. Details in Appendix A.11.

Absence and Damage Sweeps. Figure 4 draws the two sweeps on the primary cohort. Across the three cohorts, averaged along the absence sweep, the learned filling leads this work on every cohort and mean filling on two of three, by at most 0.0316, while zero-filling leads on none; along the damage sweep the gate reaches or passes this work at two of twelve settings and falls behind at the remaining ten. What the filling families buy on the task, where they lead, they pay for at the separation readout, and carrying both at once is what this work is for. Details in Appendix A.11.

5.3 Ablation Study and Analysis

Two-Coordinate Readout. Table 3 removes one component at a time at the middle working point. Collapsing the readout to one non-negative scalar per channel costs 0.4442 of separation probe, from 0.9944 to 0.5502, and buys 0.0146 of accuracy; every other variant leaves the probe at 0.9928 or above. The collapsed variant reproduces, inside this work, the trade the filling families make: a slightly better decision and a vector a probe can barely read above the random-feature floor. Its scalar is the channel's trust-weighted assertion summed over the prototypes and normalised by the largest such value over the evaluated set, where the gate reads its scalar from the channel's own dispersion. Details in Appendix A.12.

Trust Product. Figure 5 reads the variants at all three working points, to rule out that either loss belongs to the middle point alone. Holding the trust constant pins the rank agreement at exactly 0.5000 everywhere, the value of a reading with no ordering in it, while the product the full model resolves rises with the degradation, from 0.9377 to 0.9724. The first panel shows the other collapse holding as widely: the one-scalar readout reads between 0.4738 and 0.6523 over the nine cohort-by-working-point cells, against a full model that never falls below 0.9884. Neither loss is a property of the level at which the variants happen to be read.

Provenance-Typed Transport. Table 4 reports how the two masses answer to each failure at the same severity. Absence puts between 0.3651 and 0.4324 of the support mass on channels that never arrived, and damage, where every channel arrives, puts none there; damage instead lifts the conflict share to between 0.8328 and 0.8671, between 0.3117 and 0.3682 above where absence of the same

severity leaves it. A learned filling restores a comparable share, between 0.3401 and 0.3518, so the amount restored is not what tells the two failures apart. What does is the coordinate each moves: the typed transport writes restored mass into the support mass alone, and an absent channel contributes nothing to the conflict mass by construction. Details in Appendix A.11.

Algebraic Claims at Machine Precision. Table 5 reports what the two masses do when the quantities the propositions speak about are edited directly. Replacing the restored evidence by zero leaves the conflict mass bit-identical, while the conservation identity closes to 1.8×10^{-15}. The zero is not a small number read at a coarse scale: on the same instances the same replacement moves the support mass by between 1.0356 and 1.4108, fifteen orders of magnitude above the rounding floor the identity sits at. Every conflict term of Eq. 6 carries the presence factor a_m and the restored term enters Eq. 5 under $(1 - a_m)$: the separation follows from where the terms are written rather than from a quantity the fitting has to preserve.

Table 5: Largest absolute residual of each check, at absence 0.3 and damage 0.2.

Cohort	Conservation	Conflict moved	Support moved
Syn-MOSEI	1.776×10^{-15}	0	1.0356
Syn-IEMOCAP	1.776×10^{-15}	0	1.4108
Syn-ADNI	8.882×10^{-16}	0	1.4067

A Graph Estimated from Incomplete Instances. Table 21 re-estimates the compatibility graph from instances that are themselves incomplete. At an absence rate of 0.2 in the estimation set the relative deviation is between 0.0675 and 0.1688 with between 0.9167 and 0.9667 of the reference's edges surviving, and at 0.4 the deviation is between 0.1001 and 0.2218 with between 0.8917 and 0.9500 surviving. Six to thirteen of the graph's 120 edges are replaced at the higher rate, and the deviation moves with that replacement rather than staying flat. What bounds the effect is where the graph enters, the restricted operator of Eq. 12 and the reachable fraction of Eq. 13: a withdrawn edge removes a route, and an edge the estimate adds can draw only on a prototype some present channel covers, since the operator normalises over those alone. The graph the reported passes use is still estimated on complete training instances; this reading bounds what giving that up would cost.

6 Related Work

Reconstruction and Imputation for Missing Modalities. One line reconstructs what an absent channel would supply, through invariant codes (Zeng et al., 2022; Zuo et al., 2023; Li et al., 2024) or diffusion complementation (Jin et al., 2025; Zhuang et al., 2025). Both return restored content through the coordinates that held observed content. Unlike them, TwoChannel keys restored evidence to its provenance: completing an absent channel raises support and manufactures no conflict.

Uncertainty and Evidential Reliability Estimation. A second line estimates how far each channel can be trusted, as a belief mass (Huang et al., 2022; Luo et al., 2025; Pandey et al., 2026) or as a scalar that attenuates it (Meng et al., 2026; Kumar et al., 2026; Wu et al., 2025). The latter deliver one figure per channel, with no annotation of which channel is degraded, and both failures drive it toward the same value. Unlike them, TwoChannel reads a per-prototype trust from within-channel dispersion and cross-channel agreement, and carries the two failures on two coordinates.

Prototype and Graph-based Multimodal Fusion. A third line matches an instance to learned prototypes before fusion (Liu et al., 2025; Lu et al., 2026; Zhu et al., 2025) or propagates evidence over a channel graph to fill the absent one (Li et al., 2023b;a; Liang et al., 2025). Both collapse that state into a single readout, in which little separates a recovered channel from a corrupted one. Unlike them, TwoChannel propagates on a compatibility graph while keeping a two-coordinate readout, so prototype-level restoration and trust survive the fusion step.

7 Conclusion

This paper identified that incumbent fusion rules write absence and corruption onto one attenuation axis, where the margin between them narrows as damage deepens. TwoChannel keeps them apart on a support and a conflict coordinate: an absent channel contributes no conflict, and trust comes from the instance. A probe recovers the distinction at 0.9960 against mean filling's 0.4995, and complete-input accuracy reaches 0.9383 against the gate's 0.7883. Which failure occurred is a matter of where evidence is written, not of how strongly it is weighted.

REFERENCES

AmirAli Bagher Zadeh, Paul Pu Liang, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency. Multimodal language analysis in the wild: CMU-MOSEI dataset and interpretable dynamic fusion graph. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2236–2246, Melbourne, Australia, 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1208. URL https://aclanthology.org/P18-1208/.

Jindi Bao, Jianjun Qian, Mengkai Yan, and Jian Yang. Progressive representation learning for multimodal sentiment analysis with incomplete modalities. *arXiv preprint arXiv:2603.09111*, 2026.

Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N. Chang, Sungbok Lee, and Shrikanth S. Narayanan. IEMOCAP: interactive emotional dyadic motion capture database. *Language Resources and Evaluation*, 2008. doi: 10.1007/s10579-008-9076-6.

Rongfei Chen, Tingting Zhang, Xiaoyu Shen, and Wei Zhang. Evaluation before generation: A paradigm for robust multimodal sentiment analysis with missing modalities. *arXiv preprint arXiv:2604.05558*, 2026.

Taixi Chen and Nancy Guo. Learning from reliable latent prompts for visual recognition with missing modalities. *arXiv preprint arXiv:2606.30597*, 2026.

Fangze Fu, Wei Ai, Fan Yang, Yuntao Shou, Tao Meng, and Keqin Li. SDR-GNN: Spectral domain reconstruction graph neural network for incomplete multimodal learning in conversational emotion recognition. *Knowledge-Based Systems*, 2025. doi: 10.1016/j.knosys.2024.112825.

Zongbo Han, Changqing Zhang, H. Fu, and Joey Tianyi Zhou. Trusted multi-view classification. In *International Conference on Learning Representations*, 2021.

Ling Huang, Thierry Denoeux, Pierre Vera, and Su Ruan. Evidence fusion with contextual discounting for multi-modality medical image segmentation. *arXiv preprint arXiv:2206.11739*, 2022.

Shuo Jiang, Zhuwen Chen, Liaoman Xu, Yanming Zhu, Changmiao Wang, Jiong Zhang, Feiwei Qin, Yifei Chen, and Zhu Zhu. Multimodal feature prototype learning for interpretable and discriminative cancer survival prediction. *arXiv preprint arXiv:2510.06113*, 2025.

Yuehan Jin, Xiaoqing Liu, Yiyuan Yang, Zhiwen Yu, Tong Zhang, and Kaixiang Yang. Rohydr: Robust hybrid diffusion recovery for incomplete multimodal emotion recognition. *arXiv preprint arXiv:2505.17501*, 2025.

Suraj Kumar, Mohnish Raj, Soumi Chattopadhayay, Chandranath Adak, and Ayan Dutta. Adaptive modality reliability diagnosis and restoration for robust multimodal intent recognition. *arXiv preprint arXiv:2608.03475*, 2026.

Dongyuan Li, Yusong Wang, Kotaro Funakoshi, and Manabu Okumura. Joyful: Joint modality fusion and graph contrastive learning for multimoda emotion recognition. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023a. doi: 10.18653/v1/2023.emnlp-main.996.

Jiang Li, Xiaoping Wang, Guoqing Lv, and Zhigang Zeng. GraphMFT: A graph network based multimodal fusion technique for emotion recognition in conversation. *Neurocomputing*, 2023b. doi: 10.1016/j.neucom.2023.126427.

Mingcheng Li, Dingkang Yang, Yang Liu, Shunli Wang, Jiawei Chen, Shuaibing Wang, Jinjie Wei, Yue Jiang, Qingyao Xu, Xiaolu Hou, Mingyang Sun, Ziyun Qian, Dongliang Kou, and Lihua Zhang. Toward robust incomplete multimodal sentiment analysis via hierarchical representation learning. In *Advances in Neural Information Processing Systems 37*, 2024. doi: 10.52202/079017-0895.

Xiang Li, Xianfu Cheng, Dezhuang Miao, Xiaoming Zhang, and Zhoujun Li. TF-Mamba: Text-enhanced fusion mamba with missing modalities for robust multimodal sentiment analysis. In *Findings of the Association for Computational Linguistics: EMNLP*, 2025. doi: 10.18653/v1/2025.findings-emnlp.602.

Xiang Li, Xiaoming Zhang, Dezhuang Miao, Xianfu Cheng, Dawei Li, Honggui Han, and Zhoujun Li. Mixture of disentangled experts with missing modalities for robust multimodal sentiment analysis. *Information Fusion*, 2026. doi: 10.1016/j.inffus.2026.104495.

Zeyu Liang, Hailun Xia, and Naichuan Zheng. Patch as node: Human-centric graph representation learning for multimodal action recognition. *arXiv preprint arXiv:2512.21916*, 2025.

Hong Liu, Haosen Yang, Federica Eduati, Josien P. W. Pluim, and Mitko Veta. Adaptive prototype learning for multimodal cancer survival analysis. *arXiv preprint arXiv:2503.04643*, 2025.

Xiwei Liu, Yulong Li, Feilong Tang, and Imran Razzak. Delo: Dual decomposed low-rank experts collaboration for continual missing modality learning. *arXiv preprint arXiv:2603.01632*, 2026.

Jueqing Lu, Yuanyuan Qi, Xiaohao Yang, Shuaicheng Niu, Fucai Ke, Shujie Zhou, Wei Tan, Jionghao Lin, Wray L. Buntine, Hamid Rezatofighi, and Lan Du. DPL: Decoupled prototype learning for enhancing robustness of vision-language transformers to missing modalities. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2026.

Hui Luo, Jiashuang Huang, Hengrong Ju, Tianyi Zhou, and Weiping Ding. Multimodal multi-instance evidence fusion neural networks for cancer survival prediction. *Scientific Reports*, 2025. doi: 10.1038/s41598-025-93770-3.

Haoran Ma, Yinfeng Yu, and Liejun Wang. Mruf: Multi-granularity routing with uncertainty-aware fusion for robust multimodal sentiment analysis. *arXiv preprint arXiv:2607.10599*, 2026.

Chunlei Meng, Jacqueline J. Pang, Pengbin Feng, Zhenyu Yu, Chun Ouyang, and Zhongxue Gan. Rethinking modality reliability in multimodal sentiment analysis with incomplete observations. *arXiv preprint arXiv:2608.03611*, 2026.

Susanne G. Mueller, Michael W. Weiner, Leon J. Thal, Ronald C. Petersen, Clifford Jack, William Jagust, John Q. Trojanowski, Arthur W. Toga, and Laurel Beckett. The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 2005. doi: 10.1016/j.nic.2005.09.008.

Duy A. Nguyen, Quan Huu Do, Khoa D. Doan, and Minh N. Do. Are you sure? enhancing multimodal pretraining with missing modalities through uncertainty estimation. *arXiv preprint arXiv:2504.13465*, 2025.

Deep Shankar Pandey, Hyomin Choi, and Qi Yu. Generalized regularized evidential deep learning models: Theory and comprehensive evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026. doi: 10.1109/tpami.2026.3660699.

Zhenyu Wang, Weichen Cheng, Weijia Li, Junjie Mou, Zongyou Zhao, and Guoying Zhang. Urmf: Uncertainty-aware robust multimodal fusion for multimodal sarcasm detection. *arXiv preprint arXiv:2604.06728*, 2026.

Yuhua Wen, Yingying Zhou, Qifei Li, Yingming Gao, Zhengqi Wen, Jianhua Tao, and Ya Li. Midas: Mutual information disentanglement with uncertainty-aware fusion for incomplete multimodal sentiment analysis. *arXiv preprint arXiv:2608.09986*, 2026.

Han Wu, Yanming Sun, Yunhe Yang, and Derek F. Wong. Beyond simple fusion: Adaptive gated fusion for robust multimodal sentiment analysis. *arXiv preprint arXiv:2510.01677*, 2025.

Yefeng Wu. Bridging multimodal fusion and expert routing via spectral reliability descriptors for robust object detection. arXiv preprint arXiv:2606.01173, 2026.

Yucheng Xing, Hailan Mo, Zi Wang, Ling Huang, and Mengling Feng. Evidential fusion network for multimodal survival prediction under missing modalities. *arXiv preprint arXiv:2606.20757*, 2026.

Jiandian Zeng, Tianyi Liu, and Jiantao Zhou. Tag-assisted multimodal sentiment analysis under uncertain missing modalities. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2022. doi: 10.1145/3477495.3532064.

Qingyang Zhang, Haitao Wu, Changqing Zhang, Qinghua Hu, Huazhu Fu, Joey Tianyi Zhou, and Xi Peng. Provable dynamic fusion for low-quality multimodal data. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 41753–41769. PMLR, 2023. URL https://proceedings.mlr.press/v202/zhang23ar.html.

Xingyu Zhu, Shuo Wang, Beier Zhu, Miaoge Li, Yunfan Li, Junfeng Fang, Zhicai Wang, Dongsheng Wang, and Hanwang Zhang. Dynamic multimodal prototype learning in vision-language models. In *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025. doi: 10.1109/iccv51701.2025.00241.

Yinlin Zhu, Di Wu, Ziyu Han, Zekai Chen, Wang Luo, Miao Hu, and Guocong Quan. RHEA: Reliability-harmonized reconstruction and assignment for robust multimodal-attributed graph clustering. arXiv preprint arXiv:2608.00621, 2026a.

Yitong Zhu, Yuxuan Jiang, Guanxuan Jiang, Bojing Hou, Peng Yuan Zhou, Ge Lin Kan, and Yuyang Wang. Qa-moe: Towards a continuous reliability spectrum with quality-aware mixture of experts for robust multimodal sentiment analysis. *arXiv preprint arXiv:2604.05704*, 2026b.

Yan Zhuang, Minhao Liu, Yanru Zhang, Jiawen Deng, and Fuji Ren. Tmdc: A two-stage modality denoising and complementation framework for multimodal sentiment analysis with missing and noisy modalities. *arXiv preprint arXiv:2511.10325*, 2025.

Haolin Zuo, Rui Liu, Jinming Zhao, Guanglai Gao, and Haizhou Li. Exploiting modality-invariant feature for robust multimodal emotion recognition with missing modalities. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023. doi: 10.1109/icassp49357.2023.10095836.

AI USE STATEMENT

We used generative AI tools to translate the text (Claude Opus 5.0), to check grammar and spelling and polish the wording (GPT-5.6), to guide the drawing of the figures (Gemini), and to check the content (Grok). We did not use generative AI tools to modify code. We have reviewed all AI-assisted work: we checked every reference for hallucinated citations. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The details needed to reproduce the results of this paper, including the construction of the cohorts and degradations, the baselines, the metrics and the probe protocol, the dictionary, graph, trust and training settings, the forward pass with its training update, and the regeneration of the reported quantities, are disclosed in this appendix.

A APPENDIX

Table 6 lists what each part of the appendix contains.

Appendix A.0: Limitations

A damage regime that both trust factors miss. A reader who acts on the trust values needs to know how the method behaves when a channel is corrupted uniformly over its whole extent and in agreement with the other present channels, since such a corruption leaves no disagreement inside the channel and no distance from the other channels to read, and a high trust given to it would reach the decision unnoticed. This work lacks an evaluation of that regime; we did not include it in the present study, and it can be reported upon request.

Table 6: What each part of the appendix contains.

Part	Contents
A.0	Limitations
A.1	Related Work
A.2	Motivation Measurements in Full
A.3	Proofs of the Conservation and Separation Statements
A.4	Provenance-Typed Transport in Full
A.5	The Trust Product in Full
A.6	The Forward Pass and One Training Step
A.7	Cohorts and Degradations
A.8	Baselines in Full
A.9	Metrics and Probe Protocol in Full
A.10	Dictionary, Graph, Trust and Training Settings
A.11	The Main Results in Full
A.12	The Ablation Variants in Full
A.13	Compatibility-Graph Estimation
A.14	Regeneration of the Reported Quantities
A.15	Cost and Complexity
A.16	Discussion

A breakdown by how much of an instance arrived. Pooled task readings mix instances that deliver one, two or three channels, so a reader cannot tell how the reading of this method changes from the well-measured instances to the sparsely measured ones, which are the cases that matter most in deployment. This work lacks a breakdown of the task readings by the number of delivered channels; we did not include it in the present study, and it can be reported upon request.

Transfer to a new cohort. A reader who deploys the method on a new population needs to know how much of a fitted model carries over without refitting, since refitting is not always possible for a new cohort. This work lacks an evaluation that moves a model fitted on one cohort to another cohort without refitting; we did not include it in the present study, and it can be reported upon request.

Appendix A.1: Related Work

Reconstruction and Imputation for Missing Modalities. One line of work recovers an absent channel by reconstructing the evidence it would have supplied, learning tag-conditioned and modality-invariant representations that fill the empty slot before fusion (Zeng et al., 2022; Zuo et al., 2023; Li et al., 2024). Later systems add spectral-domain and hierarchical recovery that estimates the missing content from the remaining channels (Fu et al., 2025; Bao et al., 2026; Li et al., 2025). A parallel strand develops generative recovery: diffusion and denoising complementation resynthesize a degraded or absent channel and then read the completed representation (Jin et al., 2025; Zhuang et al., 2025). More recent methods examine when to trust the reconstruction itself, proposing an evaluate-before-generate stage and a mutual-information disentanglement that separates shared from channel-specific content before completion (Chen et al., 2026; Wen et al., 2026). These methods share one assumption, that a filled-in readout is as good as one whose channels were all present, so the reconstructed content re-enters through the same coordinates as observed content, and the margin a decision map downstream has for telling a recovered channel from one that was truly there narrows with it. Unlike this line of work, TwoChannel routes restored evidence on a coordinate keyed by its provenance, and completing an absent channel raises support without manufacturing the conflict signal that marks a corrupted one.

Uncertainty and Evidential Reliability Estimation. A second community estimates how far to trust each channel, developing evidential fusion with contextual discounting and multi-instance evidence networks that attach a belief mass to every channel before combining them (Huang et al., 2022; Luo et al., 2025; Pandey et al., 2026). Later work adds uncertainty-aware pretraining and evidential survival fusion that scores channel reliability from the instance itself and down-weights the unreliable one (Nguyen et al., 2025; Xing et al., 2026; Wang et al., 2026). A related strand introduces multi-granularity routing that reports a per-channel or per-token reliability and reallo-

cates fusion weight accordingly (Ma et al., 2026; Wu, 2026). In the scalar-gated family, recent methods rethink modality reliability under incomplete observation and design adaptive diagnosis-and-restoration and adaptive gated fusion that attach a single non-negative scalar to attenuate a channel (Meng et al., 2026; Kumar et al., 2026; Wu et al., 2025). The reliability score these methods attach is one figure per channel, fit end to end against task loss with no annotation of which channel is degraded. Because absence and corruption both drive that single figure toward the same value, the margin a readout carrying it leaves between a removed and a damaged channel narrows as the damage deepens. Unlike them, TwoChannel derives a per-prototype trust from within-channel residual dispersion and cross-channel agreement without degradation labels, and carries absence and corruption on two separate readout coordinates rather than on one attenuation scalar.

Prototype and Graph-based Multimodal Fusion. A third line of work represents each channel through learned prototypes, developing adaptive and decoupled prototype learning that matches an instance to a bank of class or concept anchors before fusion (Liu et al., 2025; Lu et al., 2026; Zhu et al., 2025). Later methods present interpretable feature prototypes and reliable latent prompts that recover a missing channel from the anchors the remaining channels still activate (Jiang et al., 2025; Chen & Guo, 2026; Zhu et al., 2026a). A parallel strand designs graph fusion, introducing message passing over a channel or token graph that propagates evidence across the present channels to fill and denoise the absent one (Li et al., 2023b;a; Liang et al., 2025). Recent systems learn expert routing, presenting disentangled and quality-aware mixtures and low-rank expert collaboration that select a specialized combiner for each instance (Li et al., 2026; Zhu et al., 2026b; Liu et al., 2026). These methods share a fusion step that collapses the prototype or graph state into a single readout the decision map then reads. The further that collapse is carried, the less of the separation between a recovered channel and a corrupted one survives it. Unlike this line of work, TwoChannel propagates on a compatibility graph while keeping a two-coordinate readout, so prototype-level restoration and per-prototype trust survive the fusion step instead of being averaged into one attenuated vector.

Appendix A.2: Motivation Measurements in Full

Each panel of Figure 2 is redrawn here on its own, over all three cohorts rather than the primary cohort alone. The full sweep behind panel (a) is given in Table 7 and Table 8, and those behind panels (b) and (c) in Table 9 and Table 10.

Panel (a): the separation margin. Figure 6 plots, for each one-coordinate family, the distance between a dropped channel and a damaged one in the representation that family hands downstream, divided by the distance the damage alone moved the same instance. The quantity is a ratio and not an area under a curve, its median is taken over instances, and the dashed curve is the same ratio computed on the raw input the fused representations are derived from; it is the reference the fused curves are read against rather than a competitor. The damage level is swept from 0.0500 to 0.6000 over 12 settings. Every family falls across the sweep: the three filling families lie between 1.1511 and 3.6062, the raw-input reference between 1.4557 and 6.5305, and the scalar gate below all of them at all 36 cohort-by-setting readings, reaching 0.0001 at the heaviest damage. What the panel shows is a margin that the damage itself closes; how much a probe can still recover from such a representation is a separate reading and is reported in Table 1.

Panel (b): what a filling leaves of the absence record. Figure 7 asks a linear probe which channel was never there, reading only the representation handed downstream, and compares a representation that fills nothing against one produced by a learned filling. Two absence processes are drawn and they are not interchangeable. The curves are the first process: channels are withheld independently at a rate, swept from 0.0500 to 0.6500 over 13 settings. Here the filling has most of the instance to write from, the probe falls to between 0.5065 and 0.8807, and the filling removes between 0.0997 and 0.4847 of what was readable without it. The single marker per cohort is the second process: a fixed construction in which exactly one channel of an instance is absent, and the filled value is a function of the channels that remain. There the probe reads 0.9991 and above and the filling removes at most 0.0002. The shaded region between each cohort's pair of curves is the last column of Table 9, the share of the absence record the filling removes, and it is the quantity this challenge is about; the band around the filled curve in Figure 2 is the range the filled reading takes across the three cohorts. The probe of this panel is read without the standardisation used for Table 1 and at a

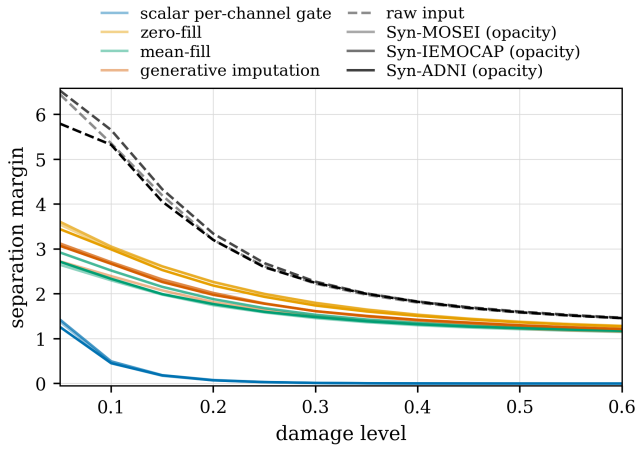

Figure 6: Separation margin against damage level, all three cohorts. Solid curves are the four one-coordinate families; the dashed curve is the same margin on the raw input.

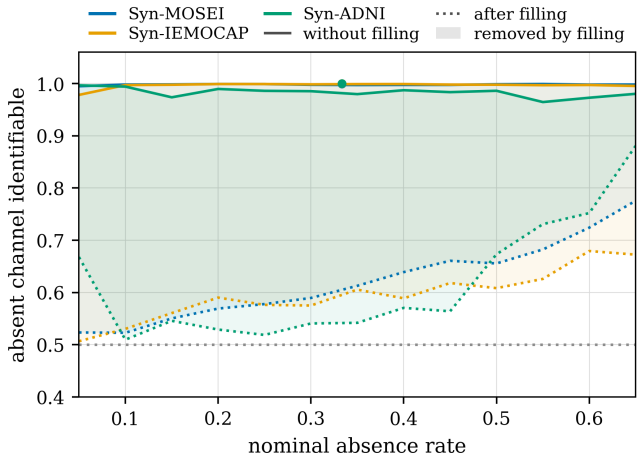

Figure 7: Absent-channel identifiability against the rate at which channels are withheld, all three cohorts. Shaded: the record a filling removes.

ridge penalty of 0.1, which makes it a different measurement from the separation probe reported in the experiments.

Panel (c): the two signals an instance carries. Figure 8 plots the two quantities that remain available when no serving-time label names the degraded channel: the spread of observable dispersion across an instance's channels, and the spread of the internal-dispersion factor those channels imply. Both are measured before any component of the method consumes them, over 21 corruption levels from 0.0000 to 0.5000. Both grow with the damage, the observable dispersion spread running from between 0.0212 and 0.0248 at no damage to between 0.4637 and 0.4683 at the highest level, and the implied trust signal from between 0.1050 and 0.1271 to between 0.2537 and 0.2771. The band around each curve in Figure 2 is the range the signal takes across the three cohorts; that range never exceeds 0.0046 for the observable dispersion, which is narrower than the line width and therefore not separately visible, against 0.0255 for the implied trust signal. The two shapes differ: the observable dispersion keeps rising across the whole sweep while the implied trust signal flattens above a corruption level of about 0.25. A weight can therefore be read from the instance, and what the instance does not supply is a statement of which reading is correct.

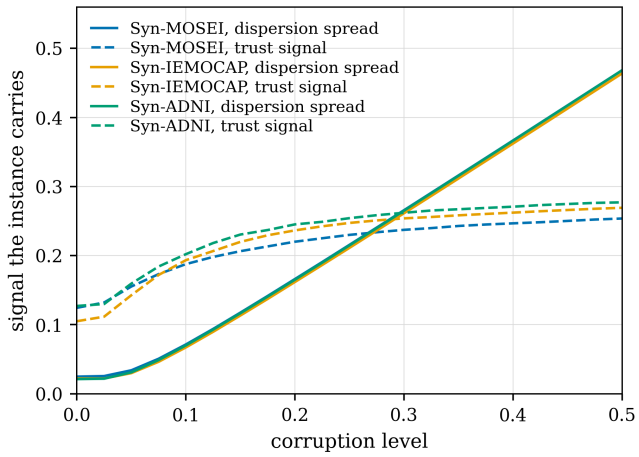

Figure 8: The two signals an instance carries against corruption level, all three cohorts. Solid: observable dispersion spread. Dashed: implied internal-dispersion factor.

Table 7: Challenge #1. Separation margin ↑ of the scalar gate against damage level, per cohort, beside the same margin on the raw input.

	Syn-MOSEI		Syn-IEMOCAP		Syn-ADNI	
Damage level	Scalar gate	Raw input	Scalar gate	Raw input	Scalar gate	Raw input
0.0500	1.3795	6.4459	1.4245	6.5305	1.2603	5.7935
0.1000	0.4570	5.3544	0.4955	5.6532	0.4538	5.3240
0.1500	0.1738	4.1957	0.1865	4.3336	0.1819	4.0572
0.2000	0.0722	3.2003	0.0783	3.3426	0.0721	3.2014
0.2500	0.0306	2.6149	0.0338	2.6851	0.0314	2.5929
0.3000	0.0127	2.2263	0.0147	2.2715	0.0135	2.2437
0.3500	0.0054	1.9783	0.0065	2.0083	0.0059	1.9994
0.4000	0.0024	1.8050	0.0028	1.8296	0.0026	1.8234
0.4500	0.0010	1.6771	0.0012	1.7014	0.0011	1.6878
0.5000	0.0004	1.5825	0.0005	1.6046	0.0005	1.5894
0.5500	0.0002	1.5123	0.0002	1.5286	0.0002	1.5154
0.6000	0.0001	1.4557	0.0001	1.4672	0.0001	1.4591

Appendix A.3: Proofs of the Conservation and Separation Statements

Each condition of the admissible class is addressed by one component and one equation. Condition (a), separation, is addressed by the two masses of Eq. 5 and Eq. 6 together with the propagation of Eq. 8, which prototype decomposition supplies. Condition (b), provenance, is met by the reachable weight of Eq. 13, which provenance-typed transport supplies. Condition (c), no annotation, is met by the trust of Eq. 16, which the trust product supplies. The training objective of Eq. 17 fits the three together under channel removal.

This section first bounds what a fusion rule carrying one weight per channel can keep, and then proves the three statements the method makes about its own masses. Each of the latter is an identity in the quantities of Eqs. 4–6 and holds for every instance, not on average; two are also checked numerically on the cohorts, and the residuals of those checks are given after the proof.

Proposition 1 (the margin a single weight leaves)**.** *Let a fusion rule read channel m through one non-negative scalar $t_m = a_m \varphi(\mathrm{disp}_m)$, with φ continuous, strictly decreasing and strictly positive, and let the observable dispersion disp_m increase in the damage level δ. Then the absent and the damaged state of that channel are separated on the coordinate by $\mu(\delta) = \varphi(\mathrm{disp}_m(\delta))$, which decreases in δ, and any Λ-Lipschitz map h of the coordinate obeys $|h(t_{\mathrm{dam}}) - h(t_{\mathrm{abs}})| \le \Lambda\, \mu(\delta)$.*

Table 8: Challenge #1, continued. Separation margin ↑ of the three filling families against damage level, per cohort.

	Syn-MOSEI			Syn-IEMOCAP			Syn-ADNI		
Damage level	Zero	Mean	Gen.	Zero	Mean	Gen.	Zero	Mean	Gen.
0.0500	3.5441	2.6478	2.7250	3.6062	2.9222	3.1227	3.4424	2.7157	3.0705
0.1000	3.0138	2.3016	2.3956	3.0548	2.5205	2.7069	2.9915	2.3356	2.6703
0.1500	2.6107	1.9830	2.0771	2.6147	2.1566	2.3211	2.5315	1.9947	2.2647
0.2000	2.2597	1.7468	1.8178	2.2663	1.8826	2.0207	2.1847	1.7726	1.9781
0.2500	1.9941	1.5856	1.6334	2.0000	1.6811	1.7872	1.9341	1.5995	1.7810
0.3000	1.8020	1.4584	1.4983	1.7972	1.5335	1.6127	1.7400	1.4820	1.6141
0.3500	1.6519	1.3676	1.3981	1.6472	1.4275	1.4960	1.6070	1.3888	1.5081
0.4000	1.5358	1.3003	1.3292	1.5334	1.3503	1.4072	1.5084	1.3215	1.4206
0.4500	1.4495	1.2499	1.2728	1.4434	1.2932	1.3401	1.4299	1.2704	1.3598
0.5000	1.3809	1.2118	1.2258	1.3713	1.2462	1.2868	1.3644	1.2350	1.3000
0.5500	1.3293	1.1786	1.1867	1.3139	1.2103	1.2474	1.3165	1.2017	1.2540
0.6000	1.2890	1.1511	1.1599	1.2706	1.1818	1.2122	1.2792	1.1750	1.2161

Table 9: Challenge #2. Absent-channel identifiability ↑ after and without a filling, and the share the filling removes, per cohort. The last row has exactly one channel absent per instance.

	Syn-MOSEI			Syn-IEMOCAP			Syn-ADNI		
Absence rate	After	Without	Removed	After	Without	Removed	After	Without	Removed
0.0500	0.5233	0.9949	0.4716	0.5065	0.9783	0.4718	0.6677	0.9967	0.3290
0.1000	0.5227	0.9981	0.4755	0.5302	0.9971	0.4669	0.5095	0.9942	0.4847
0.1500	0.5502	0.9984	0.4482	0.5604	0.9977	0.4373	0.5456	0.9737	0.4281
0.2000	0.5691	0.9992	0.4301	0.5901	0.9991	0.4090	0.5289	0.9897	0.4608
0.2500	0.5777	0.9991	0.4214	0.5762	0.9991	0.4229	0.5186	0.9861	0.4675
0.3000	0.5894	0.9976	0.4082	0.5749	0.9985	0.4235	0.5406	0.9854	0.4448
0.3500	0.6127	0.9972	0.3846	0.6054	0.9990	0.3936	0.5418	0.9798	0.4380
0.4000	0.6389	0.9974	0.3585	0.5887	0.9991	0.4104	0.5704	0.9873	0.4169
0.4500	0.6606	0.9974	0.3368	0.6180	0.9980	0.3800	0.5639	0.9836	0.4197
0.5000	0.6555	0.9987	0.3432	0.6082	0.9978	0.3896	0.6731	0.9862	0.3131
0.5500	0.6823	0.9992	0.3168	0.6259	0.9969	0.3710	0.7308	0.9646	0.2338
0.6000	0.7241	0.9980	0.2739	0.6793	0.9973	0.3180	0.7520	0.9729	0.2208
0.6500	0.7756	0.9981	0.2225	0.6724	0.9954	0.3230	0.8807	0.9804	0.0997
0.3333	0.9991	0.9993	0.0002	0.9999	1.0000	0.0001	1.0000	0.9963	-0.0037

Proof. An absent channel gives $a_m = 0$ and hence $t_m = 0$, while a damaged one gives $t_m(\delta) = \varphi(\mathrm{disp}_m(\delta)) > 0$ by strict positivity of φ, and the two states differ on the coordinate by $\mu(\delta)$. Since disp_m increases in δ and φ is strictly decreasing, μ decreases in δ. The Lipschitz bound is immediate from $|t_{\mathrm{dam}} - t_{\mathrm{abs}}| = \mu(\delta)$. □

The two states stay distinct at every finite δ, and what shrinks is the margin a decision has to read them apart by. The infimum of Eq. 2 is therefore zero for this family whenever $\mu(\delta) \to 0$.

Proposition 2 (conservation)**.** *For every prototype k and every instance, $s_k + c_k$ equals the right side of Eq. 7.*

Proof. Add Eq. 5 and Eq. 6. Grouping the terms of each present channel m,

$$a_m r_{mk} g_{mk} + a_m\big[(1 - r_{mk})g_{mk} + q_{mk}\big] = a_m\big[\big(r_{mk} + 1 - r_{mk}\big)g_{mk} + q_{mk}\big] = a_m\big(g_{mk} + q_{mk}\big),$$

in which the trust r_{mk} cancels identically. The pairwise term $\sum_{m<m'} a_m a_{m'} \chi_{mm'k}$ appears in c_k only and the restored term $\sum_m (1 - a_m) w_{mk} \hat{g}_{mk}$ in s_k only, and both survive the addition once. Summing the regrouped per-channel terms over m and appending the two gives exactly Eq. 7. □

The trust therefore partitions an observed assertion between the two coordinates and never creates or destroys any of it: what the method chooses is where the evidence sits, not how much there is.

Table 10: Challenge #3. The two signals an instance carries ↑ against corruption level, per cohort, read before any component uses them.

	Syn-MOSEI		Syn-IEMOCAP		Syn-ADNI	
Corruption level	Disp. ↑	r^{int} spread ↑	Disp. ↑	r^{int} spread ↑	Disp. ↑	r^{int} spread ↑
0.0000	0.0248	0.1244	0.0221	0.1050	0.0212	0.1271
0.0250	0.0256	0.1321	0.0221	0.1116	0.0218	0.1299
0.0500	0.0340	0.1553	0.0299	0.1427	0.0310	0.1595
0.0750	0.0507	0.1732	0.0463	0.1720	0.0488	0.1842
0.1000	0.0714	0.1877	0.0671	0.1934	0.0700	0.2021
0.1250	0.0939	0.1981	0.0897	0.2066	0.0929	0.2182
0.1500	0.1175	0.2064	0.1133	0.2199	0.1166	0.2305
0.1750	0.1416	0.2133	0.1373	0.2294	0.1407	0.2370
0.2000	0.1661	0.2201	0.1618	0.2365	0.1651	0.2452
0.2250	0.1907	0.2251	0.1864	0.2424	0.1898	0.2489
0.2500	0.2156	0.2300	0.2112	0.2473	0.2148	0.2542
0.2750	0.2406	0.2338	0.2362	0.2509	0.2398	0.2585
0.3000	0.2657	0.2373	0.2613	0.2539	0.2650	0.2620
0.3250	0.2909	0.2398	0.2864	0.2559	0.2903	0.2653
0.3500	0.3161	0.2429	0.3116	0.2583	0.3156	0.2671
0.3750	0.3414	0.2450	0.3369	0.2603	0.3409	0.2690
0.4000	0.3667	0.2467	0.3622	0.2622	0.3663	0.2708
0.4250	0.3921	0.2483	0.3875	0.2642	0.3918	0.2731
0.4500	0.4175	0.2502	0.4129	0.2660	0.4172	0.2747
0.4750	0.4429	0.2522	0.4383	0.2678	0.4427	0.2763
0.5000	0.4683	0.2537	0.4637	0.2692	0.4682	0.2771

Proposition 3 (an absent channel contributes exactly zero conflict)**.** *If $a_m = 0$ then channel m contributes 0 to c_k for every k, whatever g_{mk} and q_{mk} would have been.*

Proof. Every term of c_k in Eq. 6 carries a_m as a factor: the trusted-residual term $a_m(1 - r_{mk})g_{mk}$, the contradiction term $a_m q_{mk}$, and each pairwise term $a_m a_{m'} \chi_{mm'k}$ in which m appears. Setting $a_m = 0$ annihilates all of them. The claim is exact and not approximate, since it is the vanishing of a product and not a limit. □

Proposition 4 (restoration enters the support mass only)**.** *The restored assertions $\hat{g}_{mk}$ of Eq. 11 enter s_k and enter no term of c_k; consequently, replacing $\hat{g}$ by 0 leaves c_k unchanged and changes s_k by $\sum_m (1 - a_m) w_{mk} \hat{g}_{mk}$.*

Proof. Read Eq. 5 and Eq. 6 termwise. The symbol $\hat{g}$ occurs once, in the second sum of Eq. 5. The c_k line is built from g_{mk}, q_{mk}, r_{mk} and $\chi_{mm'k}$, and χ is a function of the observed assertions of two present channels and their cosine agreement, and no argument of any conflict term is a function of $\hat{g}$. Hence $\partial c_k / \partial \hat{g}_{mj} = 0$ identically, and the stated substitution moves s_k by the displayed amount and c_k not at all. Since the restored term is weighted by $(1 - a_m)$, only channels that did not arrive are restored, and a present but damaged channel receives nothing. □

Together Propositions 3 and 4 are the separation statement: absence adds to support and nothing else.

Two of the statements, checked on the cohorts. Propositions 2 and 4 are also evaluated numerically on the primary cohort at an absence rate of 0.3 and a corruption level of 0.2. The maximum absolute violation of Proposition 2 over all instances and prototypes is 1.8×10^{-15}, which is floating-point rounding at this magnitude. Proposition 4 is checked by replacing $\hat{g}$ with zero: the conflict mass is bit-identical, gap exactly 0.0. Table 5 carries the same residuals on each of the three cohorts.

Appendix A.4: Provenance-Typed Transport in Full

This section carries the parts of the transport component that the body states without their justification.

What a channel covers. A channel covers a prototype when its mean assertion on it reaches a fixed share of its own largest mean assertion, estimated once per pass from the complete training views. The restriction is needed because an approximate dictionary leaves a small positive coefficient on prototypes a channel never sees, and carrying those through would make the present-channel evidence positive everywhere. The denominator of that average is never zero because an instance carrying no channel is outside the problem of Eq. 1.

Why the operator normalises over asserted prototypes only. The normalisation of Eq. 12 runs over the prototypes the present channels actually assert, so the row of a prototype whose whole compatibility neighbourhood is unasserted is zero and no restored value is placed on it.

What the one-way guarantee claims. The guarantee is one of one-way injection: origin marking and completion enter the support sum, and nothing removes an assertion a present channel made. That fixes where a restoration is recorded.

Appendix A.5: The Trust Product in Full

This section carries the parts of the trust component that the body states without their justification.

Why the two trust factors are not redundant. The internal-dispersion factor is computable from a single channel and answers corruption confined to part of a channel's extent. The cross-channel agreement factor is one on single-channel instances and answers corruption uniform over a channel. Neither subsumes the other, which is why the trust is their product.

Why the agreement factor is not symmetric. The numerator of Eq. 15 counts only the excess of this channel's assertion over another present channel's, and the denominator is this channel's own assertion. A symmetric absolute difference over a symmetric sum would return the same value for both channels of a present pair, which can scale a pair together but cannot order it, and ordering is what the trust has to supply.

The resolution of the boundary. A corruption uniform across every part of a channel and also agreeing with every other present channel leaves no variation for either factor to read.

How the three loss terms are averaged. The three terms of Eq. 17 are averaged rather than summed, and over different counts: the task term over instances, the reconstruction term over instances, channels, sub-views and coordinates including the slots an absent channel leaves, and the transport term over the removed channel-prototype entries. The relative weight a given λ carries therefore moves with the share of channels removed.

Appendix A.6: The Forward Pass and One Training Step

What the framework figure draws. Figure 3 lays the pass out by component. The top strip is the instance (x, A), its encoder and the L sub-views, beside the objective of Section 3; the left block is prototype decomposition down to the two-coordinate readout, the middle block provenance-typed transport, the right block the trust product, and the bottom block end-to-end training. Each titled box inside a block carries the name of the run-in paragraph of Section 4 that defines it, and the arrows labelled (g), (g, δ), (r) and $(\hat{g}, w)$ name the quantity handed from one block to the next. The figure carries no measured quantity; the numbered equations and Algorithm 1 are where each step is defined.

Algorithm 1 evaluates the construction of Section 4 in dependency order. Stages 1–4 run unchanged at training and at serving time, and the Stage 5 update is taken at training over the pass rather than per instance. The dictionary P, the transport operators $\Theta^{(m)}$, the gains ω_m and the head h are the fitted

Algorithm 1 TwoChannel: one pass over an instance, and the training update taken over the pass.

Require: content x, presence record $A = (a_m)$; dictionary P with unit-norm columns, graph W, operators $\Theta^{(m)}$, gains ω_m, head h
Require: fixed $L, \rho, \kappa, r_{\min}, \eta, \tau, \lambda_1, \lambda_2, \epsilon$; at training, target y, removed set D, and the assertions g_m^* withheld from it
Ensure: readout (Φ_β, A) and score $h(\Phi_\beta)$; at training, updated $P, \Theta^{(m)}, \omega_m, h$ and W

Stage 1: Prototype decomposition
for each present channel m, $a_m = 1$ **do**
  encode channel m of x into L pooled sub-views $z_m^{(1)}, \dots, z_m^{(L)}$
  $u_m^{(l)} \leftarrow \arg\min_{u \geq 0} \frac{1}{2}\|z_m^{(l)} - Pu\|_2^2 + \rho\|u\|_1$ ▷ Eq. 3
  $g_{mk} \leftarrow \frac{1}{L}\sum_l u_{mk}^{(l)},\ \delta_{mk} \leftarrow \frac{1}{L}\sum_l |u_{mk}^{(l)} - g_{mk}|,\ q_{mk} \leftarrow \delta_{mk}$ ▷ Eq. 4
end for
Stage 2: Trust product
for each present channel m and prototype k **do**
  $r_{mk}^{\text{int}} \leftarrow \exp(-\kappa\,\delta_{mk}/(g_{mk} + \delta_{mk} + \epsilon))$ ▷ Eq. 14
  $r_{mk}^{\text{cross}} \leftarrow$ agreement with the other present channels, penalising excess alone, 1 if none ▷ Eq. 15
  $r_{mk} \leftarrow \max(r_{\min},\ r_{mk}^{\text{int}} r_{mk}^{\text{cross}})$ ▷ Eq. 16
end for
Stage 3: Provenance-typed transport
$\bar{g}_j \leftarrow$ trusted assertion of the present channels that cover j, over $\max(1, \sum_m a_m)$ ▷ Eq. 11
for each absent channel m, $a_m = 0$ **do**
  $M^{(m)} \leftarrow W \odot \exp(\Theta^{(m)});\ T^{(m)} \leftarrow$ rows of $M^{(m)}$ normalised on $\{j : \bar{g}_j > 0\}$, a zero-sum row left zero ▷ Eq. 12
  $\hat{g}_{mk} \leftarrow \sum_j T_{kj}^{(m)} \bar{g}_j,\ w_{mk} \leftarrow \omega_m \nu_k(A)$ with $\nu_k(A)$ the reachable fraction ▷ Eq. 11, Eq. 13
end for
Stage 4: Two masses, propagation and readout
$\sigma_{mm'} \leftarrow \cos(g_m, g_{m'}),\ \chi_{mm'k} \leftarrow \min(g_{mk}, g_{m'k})(1 - \sigma_{mm'})_+$, present pairs $m < m'$ ▷ Eq. 6
$s_k \leftarrow \sum_m a_m r_{mk} g_{mk} + \sum_m (1 - a_m) w_{mk} \hat{g}_{mk}$ ▷ Eq. 5
$c_k \leftarrow \sum_m a_m[(1 - r_{mk})g_{mk} + q_{mk}] + \sum_{m<m'} a_m a_{m'} \chi_{mm'k}$ ▷ Eq. 6
$\tilde{s} \leftarrow s + \eta W(s - c)_+,\ \tilde{c} \leftarrow c + \eta W(c - s)_+$ ▷ Eq. 8
$\Phi_\beta(x, A) \leftarrow [\tilde{s}; \tilde{c}]$; emit (Φ_β, A) and the score $h(\Phi_\beta)$ ▷ Eq. 9
Stage 5: Update at training only, taken over the pass rather than per instance
if training **then**
  $\Theta^{(m)} \leftarrow$ descent step on the transport term $\lambda_2 \sum_{m \in D} \|\hat{g}_m - g_m^*\|_2^2$ alone ▷ Eq. 17
  $P \leftarrow$ ridge-regularised fit at the current codes; $h \leftarrow$ fitted on the current readouts against y ▷ Eq. 17
  $\omega_m \leftarrow$ scale matching the restored $\hat{g}_m$ to the withheld g_m^*
  at the end of the pass, $W \leftarrow$ co-assertion of complete training instances under the current P, top-τ per row ▷ Eq. 10
end if

quantities; the temperature κ, the propagation strength η, the trust floor $r_{\min}$, the sparsity penalty ρ, the graph degree τ and the two weights λ_1, λ_2 are fixed rather than fitted. Each fitted quantity is updated by the rule its own role dictates rather than by one joint gradient step: the transport operators take descent steps on the transport term of Eq. 17 alone, the dictionary is fitted with a ridge penalty at the current codes and the head on the readouts they produce, and the gains are set outside Eq. 17 as the scale that matches the restored assertions to the withheld ones. The compatibility graph of Eq. 10 is recomputed from the current dictionary once per training pass and held fixed within that pass, and no gradient passes through its estimation into the dictionary.

Appendix A.7: Cohorts and Degradations

The three cohorts are synthetic. Syn-MOSEI, Syn-IEMOCAP and Syn-ADNI carry the channel sets of CMU-MOSEI (language, visual, acoustic) (Bagher Zadeh et al., 2018), three of IEMOCAP's streams (audio, video, transcript) (Busso et al., 2008) and ADNI (clinical assessment, MRI, FDG-PET) (Mueller et al., 2005); Syn-ADNI follows that study's normal-control and Alzheimer's-disease groups, 200 instances per group.

Each instance carries a latent non-negative prototype code drawn with a group co-activation structure, so prototypes are interdependent and the compatibility graph has something to estimate: each

prototype fires on its own with probability 0.22, and the dictionary is partitioned into groups that fire together, each group firing with probability 0.30. A channel's sensitivity, which prototypes it can see at all, is drawn independently per channel and prototype at a rate of 0.62, with the one correction that a prototype no channel would see is assigned to one at random. The sensitivities are therefore not aligned to the co-activation groups, and the prototypes an absent channel leaves uncovered are scattered rather than contiguous. The dictionary columns are drawn at random and scaled to unit norm; they are not orthogonalised, and a column keeps a non-zero inner product with its neighbours, and it is the coverage threshold of Table 11, not orthogonality, that keeps a small coefficient on a prototype a channel never sees out of $\bar{g}$. A channel is exposed as several pooled sub-views, each pooling a random part of the channel's extent, which is what makes a channel's own parts able to disagree.

A channel is made absent by removing it from the presence record and with it any content: the implementation cannot read what an absent channel would have carried. A channel is damaged by perturbing its content, in one of three regimes: confined to part of the extent, where the channel's own sub-views disagree; uniform over the extent, where they do not and only disagreement with other channels remains; and uniform while also agreeing with the other present channels, which neither factor of the trust product can see. The third regime is the boundary stated in Appendix A.0.

The damage regimes and the working points. Damage comes in three regimes: confined to part of a channel's extent, uniform over the extent, and uniform while also agreeing with the other present channels. In the first two regimes each channel is damaged independently with probability 0.5 at the stated level; in the third the stated level is applied to every channel, since a regime defined by agreement across channels is not realised if only some of them are damaged. The three working points are indexed 1, 2 and 3 in the order given in the body, so the middle working point is an absence rate of 0.3 with a damage level of 0.2.

What the fitted model is given. Two quantities read the cohort's underlying prototype code. The dictionary is initialised at those prototypes perturbed by additive noise and is then refined against the reconstruction term, which starts training near the true dictionary rather than at random. The prototype attribution offset is scored against the same latent code and is a diagnostic of the decomposition, not a decision the model makes. Neither quantity is on the decision path at serving time: nothing but the instance content and its presence record enters Algorithm 1.

Appendix A.8: Baselines in Full

Each family is re-implemented and run here on the same cohorts under the same degradations. The gated family attenuates each channel's pooled content by one non-negative scalar read from the channel's own dispersion. The filling families replace an absent channel's content by zeros, by the channel's mean, or by a linear map fitted from the present channels. The complete-modality reference never sees the degradation and is read on the undegraded view; it is a reference and not a competitor, which is why its row is flat across every sweep.

Where two families coincide because the setting removes what distinguishes them — with no channel absent there is nothing for a filling rule to fill — they are reported as one row named for that fact rather than as the same row several times.

What the transductive families see. The learned filling refits its linear map on the batch it is evaluated on, and mean-filling takes its channel means from that same batch; both therefore see the evaluation batch before predicting on it, while this work fits its dictionary, transport and head on the training split alone and reads the evaluation batch one instance at a time.

Scope. The comparison is a mechanism contrast run under one protocol. Averaged along the absence sweep within each cohort this work reads 0.8510, 0.8860 and 0.7988 against the scalar-gated family's 0.7130, 0.8593 and 0.7337, and along the damage sweep 0.8613, 0.8663 and 0.7860 against 0.7491, 0.8679 and 0.7344; these are means along a sweep within a cohort and not means across cohorts.

Appendix A.9: Metrics and Probe Protocol in Full

This section states each reading in full and how the separation probe is run.

Task readings. Accuracy and macro F1 are read on the held-out split. Expected calibration error is read at complete input after a single temperature fitted on the training-split scores.

The separation probe. A linear probe is asked whether absence or damage occurred, off what a system hands downstream, and the reading is the area under its receiver operating characteristic curve. The probe is run on matched pairs: for one instance and one channel, one version has that channel dropped and the other has it damaged, and the two differ in nothing else. It is fitted on one half of the pairs and read on the other, on a design standardised with the mean and the standard deviation of both halves together, which makes the probe transductive as well; this applies to every row alike, including this work, the raw-input ceiling and the random-feature control.

Mechanism readings. The trust rank agreement asks whether a channel's mean trust over the prototypes orders two present channels the way the damage applied to them does, a pair it does not order counting half. The prototype attribution offset is the mean over instances of one minus the cosine similarity between the propagated support mass and the instance's latent prototype code, reading how far the recovered attribution sits from the attribution that produced the instance. The restored and conflict shares are means over instances of each instance's own ratio; the remaining shares are pooled counts, and neither is a mean of per-batch rates.

What the probe is not. The diagnostics of the three challenges are three separate readings and none of them is this probe: the first is a distance ratio rather than an AUROC, and the second is an AUROC read without the standardisation used here and at a ridge penalty of 0.1.

Appendix A.10: Dictionary, Graph, Trust and Training Settings

Table 11 lists every setting the construction runs at. The dictionary is initialised near the cohort's own prototypes and refined against the reconstruction term; its columns are renormalised to unit norm after every update, as the formulation requires. The sparse programs are solved by a non-negative accelerated proximal method run for a fixed 60 iterations, with no fixed-point or residual test. The outer alternating minimisation stops once the relative decrease of the objective stays below 10^{-4} for three consecutive passes, capped at eighty. The compatibility graph is recomputed once per training pass from the current dictionary and held fixed within the pass, which is what keeps the reachable-weight expression well defined. The transport gains are updated against the withheld assertions of the channels removed on purpose. Table 16 reads what the fitted modulation moves: replacing the fitted values by zero moves the readout by between 0.0106 and 0.5564 and changes 0, 3 and 5 of the held-out decisions on the three cohorts.

Appendix A.11: The Main Results in Full

This section expands each table and figure of the main results, and states the conventions the captions do not repeat.

How every averaged entry is formed. Every averaged entry in the paper is taken over the per-cohort or per-setting readings and rounded once for printing; the per-cohort readings themselves are in the blocks released with the code. No entry is a mean of per-batch rates.

Task accuracy at complete input. Table 12 reports accuracy, macro F1, calibration and attribution offset with nothing absent and nothing damaged. With nothing absent there is nothing for a filling rule to fill, so the learned filling, zero-filling, mean-filling and the complete-modality reference are one system here and are reported as one row, plain fusion. The prototype attribution offset is defined only for a system that exposes a prototype decomposition, which is why it is empty for the comparison rows.

Table 11: Every optimization and estimation setting the released construction runs at, beyond the model-level values stated with the implementation details.

Setting	Value
Iterations of the non-negative sparse program	60
Alternating passes	at most 80, stopped once the relative decrease of the objective stays below 10^{-4} for 3 consecutive passes
Descent steps per pass on each $\Theta^{(m)}$	150
Magnitude bound on an entry of $\Theta^{(m)}$	4
Descent step size on each $\Theta^{(m)}$	1.0
Coverage threshold, as a share of a channel's largest mean assertion	0.2
Ridge penalty of the dictionary solve, the head solve and the separation probe	0.01
Ridge penalty of the filling baseline's map and of the motivation probes	0.1
Training instances	900 (Syn-MOSEI, Syn-IEMOCAP), 240 (Syn-ADNI)
Held-out instances	600 (Syn-MOSEI, Syn-IEMOCAP), 160 (Syn-ADNI)
Seed	20260820
Seed of the three diagnostic readings of the motivation	20260911

Table 12: All four task readings per cohort with no channel absent and none damaged. Accuracy ↑, macro F1 ↑, ECE ↓, attribution offset ↓. Bold marks this work.

System	Cohort	Accuracy ↑	Macro F1 ↑	ECE ↓	Attribution offset ↓
TwoChannel (ours)	Syn-MOSEI	**0.9383**	**0.9383**	**0.0187**	**0.2457**
Gated fusion, one scalar per channel	Syn-MOSEI	0.7883	0.7882	0.0440	–
Plain fusion	Syn-MOSEI	0.9383	0.9383	0.0197	–
TwoChannel (ours)	Syn-IEMOCAP	**0.9267**	**0.9266**	**0.0194**	**0.2338**
Gated fusion, one scalar per channel	Syn-IEMOCAP	0.8933	0.8932	0.0396	–
Plain fusion	Syn-IEMOCAP	0.9067	0.9065	0.0421	–
TwoChannel (ours)	Syn-ADNI	**0.8688**	**0.8676**	**0.0866**	**0.2455**
Gated fusion, one scalar per channel	Syn-ADNI	0.7937	0.7931	0.1139	–
Plain fusion	Syn-ADNI	0.8438	0.8420	0.1060	–

The two sweeps. Table 13 sweeps the absence rate with no channel damaged and Table 14 sweeps the damage level with no channel absent, the damage confined to part of a channel's extent, which is the regime the internal-dispersion factor of the trust product can read. In the damage sweep no channel is absent, so the filling families coincide and are again reported as one row. Each system is trained at the setting it is read at, so a column is a separately fitted model rather than one model read at several points.

The separation probe. Table 1 is printed to six decimals, and the probe protocol behind it is stated in Appendix A.9.

Where the evidence came from. Table 15 reads the two shares for every system, at an absence rate or a damage level of 0.4. The restored share is the share of the support mass supplied for a channel that never arrived; the conflict share is the share of the total mass sitting in the conflict coordinate. A family whose representation carries one fused signal per channel has no coordinate apart from the support in which a disagreement could sit, so its conflict column is empty rather than filled with a quantity that stands for something else, and its first column is read the same way as every other row.

The task figure and the mechanism figure. Figure 4 redraws the two sweeps on the primary cohort. In the right panel the learned filling, zero-filling and mean-filling read identically at every

Table 13: Task accuracy ↑ against the rate at which channels are absent, no channel damaged. Avg is the mean over the five rates. Bold marks this work.

System	Cohort	r=0.0	r=0.1	r=0.2	r=0.3	r=0.5	Avg
TwoChannel (ours)	Syn-MOSEI	**0.9383**	**0.8833**	**0.8467**	**0.8017**	**0.7850**	**0.8510**
Gated fusion, one scalar per channel	Syn-MOSEI	0.7883	0.7317	0.7000	0.6783	0.6667	0.7130
Generative imputation then fusion	Syn-MOSEI	0.9383	0.9000	0.8833	0.8683	0.8233	0.8826
Zero-fill then fusion	Syn-MOSEI	0.9383	0.8617	0.8050	0.7767	0.7400	0.8243
Mean-fill then fusion	Syn-MOSEI	0.9383	0.9050	0.8733	0.8533	0.8367	0.8813
Complete-modality reference	Syn-MOSEI	0.9383	0.9383	0.9383	0.9383	0.9383	0.9383
TwoChannel (ours)	Syn-IEMOCAP	**0.9267**	**0.9150**	**0.8883**	**0.8717**	**0.8283**	**0.8860**
Gated fusion, one scalar per channel	Syn-IEMOCAP	0.8933	0.8750	0.8617	0.8567	0.8100	0.8593
Generative imputation then fusion	Syn-IEMOCAP	0.9067	0.9033	0.9067	0.9050	0.8750	0.8993
Zero-fill then fusion	Syn-IEMOCAP	0.9067	0.8933	0.8767	0.8600	0.8633	0.8800
Mean-fill then fusion	Syn-IEMOCAP	0.9067	0.8967	0.8817	0.8750	0.8667	0.8854
Complete-modality reference	Syn-IEMOCAP	0.9067	0.9067	0.9067	0.9067	0.9067	0.9067
TwoChannel (ours)	Syn-ADNI	**0.8688**	**0.8438**	**0.8000**	**0.7688**	**0.7125**	**0.7988**
Gated fusion, one scalar per channel	Syn-ADNI	0.7937	0.7375	0.7125	0.7000	0.7250	0.7337
Generative imputation then fusion	Syn-ADNI	0.8438	0.8625	0.7875	0.7937	0.7500	0.8075
Zero-fill then fusion	Syn-ADNI	0.8438	0.8125	0.7688	0.7188	0.6813	0.7650
Mean-fill then fusion	Syn-ADNI	0.8438	0.8562	0.8313	0.8187	0.7312	0.8162
Complete-modality reference	Syn-ADNI	0.8438	0.8438	0.8438	0.8438	0.8438	0.8438

Table 14: Task accuracy ↑ against the level at which present channels are damaged, no channel absent. Avg is the mean over the four levels. Bold marks this work.

System	Cohort	ℓ=0.0	ℓ=0.1	ℓ=0.2	ℓ=0.4	Avg
TwoChannel (ours)	Syn-MOSEI	**0.9383**	**0.8917**	**0.8300**	**0.7850**	**0.8613**
Gated fusion, one scalar per channel	Syn-MOSEI	0.7883	0.7567	0.7233	0.7283	0.7491
Any filling then fusion (all three identical)	Syn-MOSEI	0.9383	0.9167	0.8867	0.8317	0.8933
Complete-modality reference	Syn-MOSEI	0.9383	0.9383	0.9383	0.9383	0.9383
TwoChannel (ours)	Syn-IEMOCAP	**0.9267**	**0.9167**	**0.8467**	**0.7750**	**0.8663**
Gated fusion, one scalar per channel	Syn-IEMOCAP	0.8933	0.8833	0.8600	0.8350	0.8679
Any filling then fusion (all three identical)	Syn-IEMOCAP	0.9067	0.8967	0.8683	0.8300	0.8754
Complete-modality reference	Syn-IEMOCAP	0.9067	0.9067	0.9067	0.9067	0.9067
TwoChannel (ours)	Syn-ADNI	**0.8688**	**0.7750**	**0.7812**	**0.7188**	**0.7860**
Gated fusion, one scalar per channel	Syn-ADNI	0.7937	0.7438	0.7125	0.6875	0.7344
Any filling then fusion (all three identical)	Syn-ADNI	0.8438	0.8313	0.8250	0.7812	0.8203
Complete-modality reference	Syn-ADNI	0.8438	0.8438	0.8438	0.8438	0.8438

level, so their three curves coincide and only one is visible. Figure 9 redraws the separation probe per cohort in panel (a) and the restored share under absence and under damage in panel (b), panel (a) from the per-cohort readings and panel (b) as the mean over the three cohorts.

Appendix A.12: The Ablation Variants in Full

The body ablation table and figure. Table 3 averages the three cohorts at the middle working point, one row per variant, and every entry there is the mean of the per-cohort rows of this section. The readings the variants are built to move are the separation probe and the trust rank agreement, which is why the body paragraph adjudicates those two. Figure 5 draws the two mechanism readings against the working point instead, each curve the mean over the three cohorts with the cohort range shaded, so a variant that moves a reading is separated from the full model at every level rather than at the one the table prints.

Table 17, Table 18, Table 19 and Table 20 read the task accuracy, the separation probe, the trust rank agreement and the reachable-slot share of the full model and of each variant at three working points on each of the three cohorts. The variants are the readout collapsed to one non-negative scalar per channel, imputation in place of the provenance-typed transport, constant trust in place

Table 15: Both shares for every system, cohort and condition. Bold marks this work.

System	Cohort	Condition	Restored share of support	Conflict share of total
TwoChannel (ours)	Syn-MOSEI	absence	**0.3651**	**0.5554**
TwoChannel (ours)	Syn-MOSEI	damage	**0.0000**	**0.8671**
Gated fusion, one scalar per channel	Syn-MOSEI	absence	0.0000	–
Gated fusion, one scalar per channel	Syn-MOSEI	damage	0.0000	–
Generative imputation then fusion	Syn-MOSEI	absence	0.3401	–
Generative imputation then fusion	Syn-MOSEI	damage	0.0000	–
TwoChannel (ours)	Syn-IEMOCAP	absence	**0.4272**	**0.4646**
TwoChannel (ours)	Syn-IEMOCAP	damage	**0.0000**	**0.8328**
Gated fusion, one scalar per channel	Syn-IEMOCAP	absence	0.0000	–
Gated fusion, one scalar per channel	Syn-IEMOCAP	damage	0.0000	–
Generative imputation then fusion	Syn-IEMOCAP	absence	0.3446	–
Generative imputation then fusion	Syn-IEMOCAP	damage	0.0000	–
TwoChannel (ours)	Syn-ADNI	absence	**0.4324**	**0.4937**
TwoChannel (ours)	Syn-ADNI	damage	**0.0000**	**0.8414**
Gated fusion, one scalar per channel	Syn-ADNI	absence	0.0000	–
Gated fusion, one scalar per channel	Syn-ADNI	damage	0.0000	–
Generative imputation then fusion	Syn-ADNI	absence	0.3518	–
Generative imputation then fusion	Syn-ADNI	damage	0.0000	–

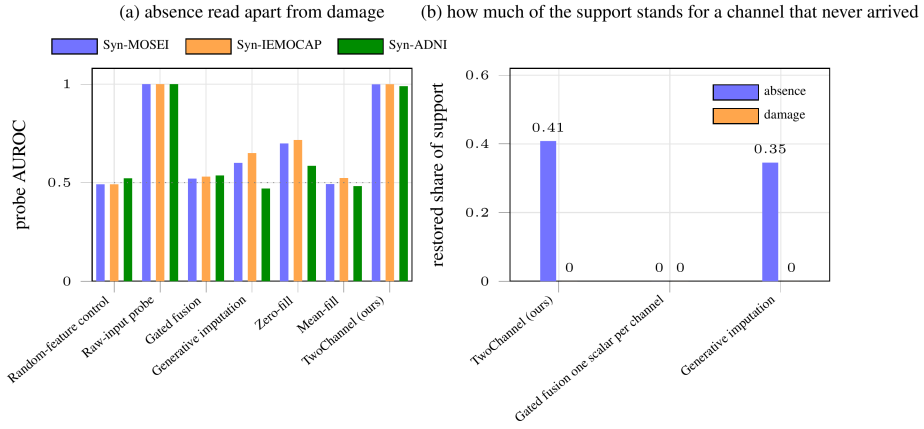

Figure 9: (a) Separation probe ↑ per cohort, against the dotted line at 0.5000. (b) Restored share of the support mass under absence and under damage, mean over the three cohorts.

of the trust product, the graph propagation removed, and residual-only trust with the cross-channel factor removed.

What imputation in place of provenance-typed transport removes. The transport operator of Eq. 12 carries a per-channel modulation $\Theta^{(m)}$ of the compatibility graph, fitted against the transport term of Eq. 17 on the instances where that channel is absent, and the operator differs by channel. This variant removes the provenance typing around it: the restored entry is given full weight in place of the reachable fraction of Eq. 13 and is admitted to the conflict coordinate. Every value in this paragraph is a mean over all nine working points of the three cohorts. That variant therefore pins the reachable-slot share at 1.0000 where the full model reads 0.9968, and reads 0.7632 on task accuracy, 0.9971 on the separation probe and 0.9611 on the trust rank agreement against the full model's 0.7610, 0.9957 and 0.9570.

Appendix A.13: Compatibility-Graph Estimation

Figure 10 plots the readings of Table 21. The deviation grows with the absence rate of the estimation set on Syn-MOSEI and Syn-ADNI, from 0.0675 and 0.1688 at an absence rate of 0.2 to 0.1153 and

Table 16: The transport modulation at the middle working point: largest fitted entry, readout movement under replacement by zero, decisions changed.

Cohort	Largest fitted entry	Readout movement when zeroed	Decisions changed
Syn-MOSEI	0.0458	0.0106	0
Syn-IEMOCAP	1.2381	0.5564	3
Syn-ADNI	0.9339	0.2771	5

Table 17: Task accuracy ↑ of the full model and of each variant at the three working points (WP) per cohort.

Cohort	WP	Full	Collapsed	Imputation	Const. trust	No graph	Residual-only
Syn-MOSEI	1	0.8500	0.8017	0.8350	0.8733	0.8533	0.8383
	2	0.7100	0.7750	0.7300	0.8067	0.7567	0.7717
	3	0.6900	0.7317	0.7100	0.7317	0.7150	0.7150
Syn-IEMOCAP	1	0.9000	0.8767	0.8900	0.8850	0.8967	0.8817
	2	0.8183	0.8283	0.8167	0.8067	0.8183	0.8067
	3	0.7433	0.7767	0.7367	0.7233	0.7517	0.7617
Syn-ADNI	1	0.7625	0.7500	0.7750	0.8375	0.7562	0.7188
	2	0.7562	0.7250	0.7250	0.7562	0.7375	0.7750
	3	0.6188	0.6625	0.6500	0.7000	0.6438	0.6312

0.2218 at 0.4, while on Syn-IEMOCAP it falls from 0.1073 to 0.1001. The graph the transport rests on degrades rather than fails when the instances available to estimate it are themselves incomplete.

Appendix A.14: Regeneration of the Reported Quantities

Table 22 regenerates six released quantities per cohort in a separate process, once with the released seed and once with that seed moved by one. With the released seed the largest and the mean absolute gap are 0.0000 on all three cohorts, while moving the seed by one changes the quantities by at most 446.0643, 546.0359 and 304.8283 on Syn-MOSEI, Syn-IEMOCAP and Syn-ADNI, against largest quantities of 20935.4200, 18517.9800 and 4929.2200.

Appendix A.15: Cost and Complexity

Where the cost goes. What the two coordinates add over a gated baseline is an order of growth per instance, not a second pass: the sub-view decomposition costs $O(LKd)$ per channel and proximal step, the graph propagation $O(K\tau)$ on the sparsified graph, and the restricted transport $O(K\tau)$ per absent channel. All three are linear in the dictionary size K and in the graph degree τ, and none of them grows with the number of damaged channels. The transport term is paid only for channels that are absent, so an instance that arrives complete pays for the decomposition and one propagation and not for transport. The three terms are the sub-view decomposition of Eq. 3, the graph propagation of Eq. 8 and the restricted transport of Eq. 11, and the sparsified graph is what keeps the last two at $O(K\tau)$ rather than $O(K^2)$.

Appendix A.16: Discussion

Read together, the readings describe a single requirement: that a fusion rule record what a piece of evidence is and where it came from, not only how much to believe it. Separating a dropped channel from a corrupted one is then a property of where evidence is written into the readout, not of how strongly a channel is weighted. A scalar per-channel weight drives absence and present-but-degraded along the same scalar path, and the margin a probe reading the fused vector has to work with narrows as the damage deepens; the separation the readout keeps, the origin gate on restored evidence, and the floored trust product are facets of that commitment rather than unrelated additions.

Table 18: Separation probe ↑ of the full model and of each variant at the three working points (WP) per cohort.

Cohort	WP	Full	Collapsed	Imputation	Const. trust	No graph	Residual-only
Syn-MOSEI	1	0.9995	0.5999	0.9998	0.9999	0.9994	0.9998
	2	0.9942	0.5315	0.9990	1.0000	0.9987	1.0000
	3	0.9942	0.6523	0.9996	1.0000	0.9985	0.9999
Syn-IEMOCAP	1	0.9988	0.6080	0.9993	0.9997	0.9994	0.9982
	2	0.9993	0.6082	0.9995	1.0000	0.9997	1.0000
	3	0.9989	0.5693	0.9989	1.0000	0.9993	1.0000
Syn-ADNI	1	0.9983	0.4738	0.9984	0.9877	0.9931	0.9830
	2	0.9897	0.5109	0.9898	0.9908	0.9898	0.9784
	3	0.9884	0.5169	0.9897	0.9888	0.9933	0.9830

Table 19: Trust rank agreement ↑ of the full model and of each variant at the three working points (WP) per cohort.

Cohort	WP	Full	Collapsed	Imputation	Const. trust	No graph	Residual-only
Syn-MOSEI	1	0.9405	0.8396	0.9405	0.5000	0.9419	0.9488
	2	0.9389	0.9751	0.9751	0.5000	0.9751	0.9683
	3	0.9737	0.9737	0.9868	0.5000	0.9912	0.9781
Syn-IEMOCAP	1	0.9405	0.9308	0.9281	0.5000	0.9405	0.9322
	2	0.9729	0.9729	0.9729	0.5000	0.9729	0.9774
	3	0.9781	0.9781	0.9781	0.5000	0.9781	0.9825
Syn-ADNI	1	0.9321	0.9321	0.9321	0.5000	0.8210	0.8827
	2	0.9712	0.9712	0.9712	0.5000	0.9808	0.9615
	3	0.9655	0.9828	0.9655	0.5000	0.9655	0.9828

Carrying a support mass and a conflict mass per prototype gives the readout the coordinate a scalar weight lacks. An absent channel contributes no conflict, because every conflict term carries the presence record, while a present-but-degraded channel contributes its distrusted assertion and its contradiction. The variant that replaces the two masses by the pooled channel content, gated by one non-negative scalar per channel, the channel's trust-weighted assertion summed over the prototypes and divided by the largest such value over all channels and evaluated instances, has its probe fall to 0.5502, within 0.0483 of the random-feature control; the masses stay legible only because they are never renormalized back to a single total.

Filling an absent channel through the observed coordinates re-creates the collapse, while provenance-typed transport does not. Origin governs where restored evidence acts: a reconstructed assertion can raise support and can never manufacture conflict, and the gate that makes its contribution vanish for every present channel is what keeps the distinction the finer decomposition was introduced to hold. The variant that replaces transport with imputation returns restored content through the same entries as observed content and declines no slot at all, pinning the reachable-slot share at 1.0000 where the full model reads 0.9967 at the middle working point. The typing is what makes the one-sidedness hold at every damage level rather than at the levels a table happens to sample.

A reliability value fit against task loss alone falls back to a channel-level number, the resolution a scalar attenuation already reached. With nothing labelling which channel was degraded, per-prototype resolution has to be read from what the instance itself carries: the disagreement among a channel's own sub-views, and its distance from the other present channels. The variant that fixes trust to a constant loses the ordering against applied corruption, falling to 0.5000 where the full model reads 0.9610, which places that ordering in the estimator rather than the task head. The two factors of the product divide the regimes rather than the readings: removing the cross-channel factor leaves the ordering at 0.9691 on the regime the internal-dispersion factor covers, and the second factor is the only one of the two that can read a damage uniform over a channel's extent.

Table 20: Reachable-slot share of the full model and of each variant at the three working points (WP) per cohort.

Cohort	WP	Full	Collapsed	Imputation	Const. trust	No graph	Residual-only
Syn-MOSEI	1	0.9990	0.9979	1.0000	0.9975	0.9989	0.9990
	2	0.9917	0.9967	1.0000	0.9924	0.9967	0.9967
	3	0.9851	0.9851	1.0000	0.9851	0.9947	0.9947
Syn-IEMOCAP	1	1.0000	0.9999	1.0000	0.9999	1.0000	1.0000
	2	0.9987	1.0000	1.0000	0.9999	0.9987	0.9999
	3	0.9998	0.9944	1.0000	0.9944	1.0000	0.9998
Syn-ADNI	1	0.9992	1.0000	1.0000	0.9997	0.9997	0.9997
	2	0.9997	1.0000	1.0000	0.9997	0.9997	0.9997
	3	0.9984	0.9995	1.0000	0.9984	0.9984	0.9984

Table 21: The compatibility graph estimated at an estimation-set absence rate of 0.2 or 0.4, against the graph the complete set gives; at a rate of 0 the two coincide.

	Relative deviation ↓		Edge overlap ↑	
Cohort	0.2	0.4	0.2	0.4
Syn-MOSEI	0.0675	0.1153	0.9667	0.9417
Syn-IEMOCAP	0.1073	0.1001	0.9417	0.9500
Syn-ADNI	0.1688	0.2218	0.9167	0.8917

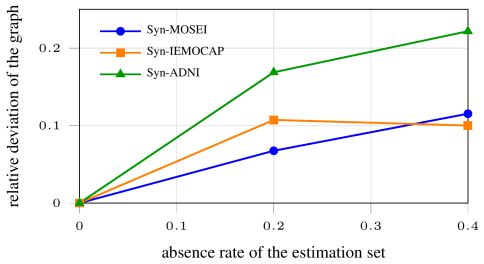

Figure 10: Relative deviation of the compatibility graph from the graph estimated on complete instances, against the absence rate of the estimation set.

Table 22: Regeneration of six released quantities per cohort in a separate process, reported as the largest absolute gap ↓.

Cohort	Comparison	Max abs gap	Mean abs gap	Largest quantity
Syn-MOSEI	across processes, released seed	0.0000	0.0000	20935.4200
Syn-MOSEI	released seed moved by one	446.0643	100.5767	20935.4200
Syn-IEMOCAP	across processes, released seed	0.0000	0.0000	18517.9800
Syn-IEMOCAP	released seed moved by one	546.0359	132.4530	18517.9800
Syn-ADNI	across processes, released seed	0.0000	0.0000	4929.2200
Syn-ADNI	released seed moved by one	304.8283	56.3451	4929.2200

A reading the design leaves open concerns channels whose encoder emits a single pooled vector, where the internal-dispersion factor is empty by construction and the trust rests on cross-channel agreement alone. A direct next experiment partitions such an encoder's feature coordinates in place of its token or frame positions and asks whether the dispersion read from that partition recovers the separation that the position-based partition cannot reach; a flat reading fixes the case as a genuine boundary, while a reading that tracks the multi-channel case extends the mechanism to cohorts the current construction excludes.

Several directions extend this work. First, a compatibility-graph estimator that stays accurate when it is fit on incomplete rather than complete training instances would remove the clean-data requirement the current construction rests on, and comparing such an estimator against the complete-instance graph on the primary cohort would test whether the estimation bias can be closed rather than only bounded. Second, the separated readout supplies an abstention and escalation signal distinct from a scalar confidence, and building a decision policy on it and evaluating that policy against a scalar-confidence baseline on a serving stream would test whether the separation lowers downstream risk rather than only sharpening diagnosis. Third, reading the fused output above a second encoder family on a real, non-synthetic clinical multimodal cohort would test whether the mechanism sits

above the encoders it consumes, as the account here claims. Each direction names the cohort and the comparison it needs.