THE RELATIVITY OF CAUSAL KNOWLEDGE: BACKBONE IDENTIFIABILITY UNDER PRIVATE KNOWLEDGE

Anonymous authors
Paper under double-blind review

ABSTRACT

Complex causal models may contain far more detail than a shared task requires. A Markov kernel can retain the outcome distributions for selected interventions while omitting other variables and mechanisms. The *Relativity of Causal Knowledge* (RCK) framework organises communication of such distilled causal quantities along network paths, using shared, interventionally consistent abstractions at each step. This offers the potential to reduce communication demands and disclosure of models and data. We examine an identification requirement for realising that potential. For two causes of a common outcome, we establish local non-identifiability for general kernels under explicit conditions: complete separate intervention reports can leave the joint response undetermined. Known identification obstacles therefore persist in the abstract quantities, even when RCK transport is well defined. Under additive separability, identified response functions determine mean changes; the noise distribution additionally determines the full kernel. We thereby take a step towards operational RCK, highlighting the need to connect its communication framework with causal identification and distributed inference.

1 INTRODUCTION

Causal knowledge supports reasoning about interventional and counterfactual scenarios and allows findings from different studies to inform one another (Peters et al., 2017). *Structural Causal Models* (SCMs) encode this knowledge through variables, causal mechanisms, and assumptions about unobserved factors (Pearl, 2009). Researchers may use complex SCMs that differ in variables, mechanisms and time scales, yet need only selected intervention responses for a common task. A Markov kernel collects these responses as outcome distributions, leaving other variables and mechanisms local. The aim is to share only what the causal question requires. The *Relativity of Causal Knowledge* (RCK) framework (D'Acunto & Battiloro, 2025) organises communication of these quantities along network paths, with a shared, interventionally consistent abstraction, or *backbone*, at each step. This could reduce communication demands and disclosure while allowing a recipient to reuse another agent's causal knowledge.

Example 1 (Education value-added: communicating causal responses)**.** Consider a policy-maker asking whether better teachers can compensate for childhood neighbourhood disadvantage, and if so by how much. Such information could guide support to help disadvantaged pupils catch up with their peers, as in England's Pupil Premium (Department for Education, 2026b). Adult earnings X provide a common outcome: teacher value-added studies estimate long-run effects of teacher quality (Chetty et al., 2014a;b), while neighbourhood studies estimate effects of childhood environment (Chetty et al., 2016; Chetty & Hendren, 2018). The teacher SCM may track classroom assignments and learning, with family background confounding assignment and earnings; the neighbourhood SCM may track residential histories, moves and family circumstances. Suppose the researchers work on separate authorised projects with confidential administrative records, sharing only outputs checked for release, as in secure research services (Office for National Statistics, 2025). For this collaboration, only interventional earnings distributions are released; records and full SCMs remain local. Given a shared, interventionally consistent backbone for the earnings response, the teacher researcher sends a kernel giving the earnings distribution under each teacher-quality intervention. The neighbourhood researcher knows responses to residential interventions but lacks the teacher-intervention

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.

responses; receiving the kernel makes these available without reconstructing classroom mechanisms or confounding adjustment.

Other settings make RCK's potential communication benefit equally concrete. An education model may send test-score distributions under teaching interventions to a labour-market model, or a pharmacology model may send biomarker distributions under drug doses to a clinical model. In a distributed environmental sensor network, local models may include device calibration, sampling times and site conditions; a shared kernel could retain regional pollutant distributions under emissions interventions for use by a health model. Heterogeneous measurements and communication constraints make this reduction especially useful in sensor networks (D'Acunto & Battiloro, 2025, Sec. 5).

In RCK, each network edge carries a shared abstraction and maps for this transfer (D'Acunto & Battiloro, 2025, Sec. 4, Defs. 8–11). *Interventional consistency* requires that intervening locally and translating the outcome give the same law as the corresponding abstract intervention. We ask: **When do private causal reports determine the shared backbone response needed for RCK communication?** We focus on the backbone's intervention-response kernel: the common causal description that the recipient needs for the shared task, while the full local models remain private.

Example 2 (Undetermined response on a simple backbone)**.** Return to Example 1. Consider a simple backbone retaining teacher-quality exposure Z_ρ, neighbourhood exposure Z_σ and adult earnings X in a common target population; classroom and residential mechanisms remain local. For a mean-level illustration at this interface, take independent, centred scalar exposures, with $Z_\sigma < 0$ denoting disadvantage, and independent mean-zero noise:

$$X_\gamma = \alpha Z_\rho + \beta Z_\sigma + \gamma Z_\rho Z_\sigma + \varepsilon.$$

For fixed slopes α, β and noise law, each γ induces a candidate response kernel on this same backbone. Averaging over the other cause gives the same private *mean* reports for every γ:

$$\mathbb{E}[X_\gamma \mid \mathrm{do}(Z_\rho = z_\rho)] = \alpha z_\rho, \qquad \mathbb{E}[X_\gamma \mid \mathrm{do}(Z_\sigma = z_\sigma)] = \beta z_\sigma.$$

The joint interventional mean, however, is $m_\gamma(z_\rho, z_\sigma) = \alpha z_\rho + \beta z_\sigma + \gamma z_\rho z_\sigma$. A unit increase in teacher quality therefore changes expected earnings by $\alpha + \gamma z_\sigma$: it has a larger effect in disadvantaged neighbourhoods when $\gamma < 0$, and in advantaged neighbourhoods when $\gamma > 0$. Both possibilities, as well as additivity ($\gamma = 0$), fit the same private means but give different responses at the communication interface. Because the teacher report averages across neighbourhoods, these means do not identify the backbone's teacher response at the disadvantaged exposure, and hence the compensating intervention.

Here the familiar interaction problem from epidemiology and experimental design (Czitrom, 1999; Greenland, 2009; Wu & Hamada, 2009; Knol & VanderWeele, 2012) leaves the causal response at the shared communication interface undetermined. Example 2 is a mean-level illustration: its full single-cause distributions need not agree as γ varies. Distributional non-identification from separate interventions also has direct precedents (Saengkyongam & Silva, 2020); a single analyst with the same reports faces the same identification problem.

We take a step towards operationalising RCK by formalising a distributed setting in which the agents' private interventional distributions provide the evidence for identifying the response kernel on their shared backbone. We contribute **(1)** a local non-identification theorem for the response represented on the RCK backbone. Around every compatible kernel satisfying explicit non-degeneracy and overlap conditions, Theorem 1 constructs arbitrarily small perturbations that preserve both *entire report distributions* while changing some joint responses. This general construction shows that restricting attention to nearby candidate responses does not resolve the ambiguity. The scalar examples and concrete witnesses illustrate the argument; the theorem concerns general kernels on measurable spaces. Even when interventionally consistent translations are available, the reports may therefore leave several possible responses at the shared backbone. We also give **(2)** sufficient recovery conditions for using RCK with these private reports. Under additive separability, identified causal response functions fix the joint response surface; fixing, identifying, or communicating the disturbance law additionally fixes the full kernel. Observational residual summaries remain insufficient. These conditions specify what causal information suffices for a recipient's calculation while detailed local models remain private. Identification is task-specific: distinct kernels can agree on a policy answer.

2 Related Work

Our question is whether private local interventional distributions identify the joint-response kernel on the abstract RCK backbone. The closest connections are causal-abstraction learning, distributed inference, causal marginal problems and joint-effect identification.

RCK, abstraction, and compositionality. RCK (D'Acunto & Battiloro, 2025) organises causal communication through interventionally consistent shared abstractions (see Section 3 and Figure 1). The semantic embedding principle (D'Acunto et al., 2025) and broader abstraction formalisms (Beckers & Halpern, 2019; Rischel & Weichwald, 2021; Massidda et al., 2023) relate models at different levels of description. Recent causal abstraction networks add learning procedures and results on knowledge diffusion, including methods for Gaussian mixtures (D'Acunto et al., 2026). These contributions address how local models are related and how knowledge moves between them. The independent-causal-mechanisms perspective (Schölkopf et al., 2012; Parascandolo et al., 2018) motivates modular causal descriptions, and Friend et al. (2025) develop compositional identification under imperfect or disturbing observations. Modularity alone does not imply additive separability at a common outcome. Our question is whether the agents' private local interventional distributions identify the backbone's joint response, even when suitable translation maps are supplied. This differs from learning or selecting an abstraction map.

Study combination, transport, and marginal compatibility. Transportability and data fusion study when causal information from different environments or studies can be combined (Pearl & Bareinboim, 2014; Bareinboim & Pearl, 2016); related work studies causal invariance across environments (Rojas-Carulla et al., 2018). Latent-variable compatibility imposes constraints on observable distributions (Tian & Pearl, 2002). The structural causal marginal problem instead determines the set of joint SCMs counterfactually consistent with overlapping marginal SCMs (Gresele et al., 2022). In categorical models, their response-function formulation also permits bounds on linear causal queries through linear programming. Our inputs are reduced intervention-response kernels rather than marginal SCMs, and we work directly on general measurable spaces. The local construction exhibits distinct joint responses close to a compatible baseline under explicit overlap conditions; it neither enumerates full SCMs nor computes sharp bounds over the entire compatible set. A particular policy query may still be identified even when the full kernel is not.

Federated methods also keep records local: Xiong et al. (2023) aggregate propensity-score-based summaries to estimate average effects of the same treatment across heterogeneous sites, under unconfoundedness and overlap assumptions. FedCDH (Li et al., 2024) uses summary statistics for conditional-independence tests and causal orientation. In our setting, the agents' initially private reports concern different causes of a shared outcome, each averaging over the other agent's cause. We take these response kernels as known exactly and ask whether they identify the response kernel required on the shared RCK backbone, including responses to interventions on both causes.

Joint effects from separate experiments. The failure of marginal effects to determine joint effects is established. Saengkyongam & Silva (2020) give a direct non-identification example in which observational and full single-variable interventional distributions agree while a joint intervention differs. Their positive result uses continuous additive-Gaussian-noise SCMs and richer observational and experimental distributions; it allows interactions between causes, unlike our outcome-separability assumption. General identification from surrogate experiments (Lee et al., 2020) and joint-effect recovery by combining experiments (Jung et al., 2023) provide further positive results using graphical knowledge and richer experimental distributions. These precedents establish the negative phenomenon and the value of structural restrictions. Our contribution is a local construction through every compatible baseline satisfying the explicit and intentionally weak assumptions (A1)–(A3) on general measurable spaces. It locates the identification problem in recovering the shared backbone response from private local interventional distributions.

Aggregating expert causal judgements. Bradley et al. (2014) study aggregation of causal-network judgements and probabilities over a common variable set. Alrajeh et al. (2018) take experts' SCMs as inputs and define compatibility and dominance conditions for merging them; Friedenberg & Halpern (2018) extend this programme to different focus areas using a *can-explain* relation. We assume access to causal reports, without the experts' SCMs or focus sets, and ask whether those reports determine the joint-intervention kernel needed for RCK communication.

Appendix E provides further discussion of these connections and the scope of our results.

3 BACKGROUND

A Structural Causal Model (SCM) M specifies (Peters et al., 2017): endogenous variables $V = (X_1, \dots, X_d)$ with state spaces $\mathcal{X}_1, \dots, \mathcal{X}_d$; exogenous variables $U = (U_1, \dots, U_d)$ with state spaces $\mathcal{U}_1, \dots, \mathcal{U}_d$ and joint law P_U; a directed acyclic graph $G = (V, E)$ whose edges represent direct causal relations; and measurable structural assignments

$$X_j := f_j(X_{\mathrm{Pa}(j)}, U_j), \qquad j = 1, \dots, d. \tag{1}$$

Here $\mathrm{Pa}(j)$ indexes the parents of X_j in G, and $X_{\mathrm{Pa}(j)}$ is their vector of values. The exogenous variables represent background factors; their joint law need not factorise, allowing unobserved common causes. Acyclicity makes the endogenous values uniquely determined by U. State spaces are measurable spaces, and products carry the product σ-algebra (Kallenberg, 2002, Chs. 1–2). We suppress these σ-algebras: an event $H \subseteq \mathcal{X}$ is always measurable. We write P_Y for the law of a random variable Y and $\perp\!\!\!\perp$ for statistical independence.

An intervention $\mathrm{do}(X_I = x_I)$, with $I \subseteq \{1, \dots, d\}$, replaces the assignments for $i \in I$ by constants $X_i := x_i$, leaving the other assignments and P_U unchanged (Pearl, 2009; Peters et al., 2017). Solving these modified assignments with $U \sim P_U$ induces the interventional law $P_M(\cdot \mid \mathrm{do}(X_I = x_I))$. We omit M when the SCM is clear. For an outcome X, we write $P(X \in H \mid \mathrm{do}(Z = z))$ for event probabilities and $\mathbb{E}[X \mid \mathrm{do}(Z = z)]$ for its expectation when X is real-valued and integrable under this law.

We detail only the components of RCK needed for our identification problem. RCK's *measure-theoretic SCMs* link exogenous and endogenous probability spaces through measurable functions (D'Acunto & Battiloro, 2025, Def. 1). The *causal knowledge* of an SCM includes that SCM and its intervened states (D'Acunto & Battiloro, 2025, Def. 4). We collect the resulting outcome laws in a Markov kernel, with intervention value z as input and outcome law $K(\cdot \mid z)$ as output, without specifying the full SCM. Formally, a Markov (probability) kernel $K : \mathcal{Z} \rightsquigarrow \mathcal{X}$ assigns each $z \in \mathcal{Z}$ a probability measure $K(\cdot \mid z)$, with $z \mapsto K(H \mid z)$ measurable for every outcome event H (Bertsekas & Shreve, 1978, Ch. 7). Here $K(H \mid z) = P(X \in H \mid \mathrm{do}(Z = z))$. The notation $K(dx \mid z)$ denotes its outcome measure, without assuming a density. This formulation covers discrete, continuous and mixed outcomes, as well as intervention trajectories, without imposing a shared parametric SCM.

In an RCK network, agents ρ, σ carry local measure-theoretic SCMs M^ρ, M^σ; their edge τ carries a shared abstract SCM M^τ, the *backbone*. The edge connects agents, not causal variables. Abstraction maps relate each local SCM to M^τ, preserving intervention responses (D'Acunto & Battiloro, 2025, Defs. 5–6). In RCK's notation, $CK(M)$ is a space of probability measures representing the SCM, and $\chi^\rho \in CK(M^\rho)$ is a local value (D'Acunto & Battiloro, 2025, Theorem 3, Sec. 4). When χ^ρ is induced by $\mathrm{do}(Z = z)$, its law on X is $K(\cdot \mid z)$. A kernel alone does not specify the abstraction maps. RCK transport requires a shared interventionally consistent abstraction and compatible maps to align the agents' causal representations (D'Acunto & Battiloro, 2025, Sec. 4, App. E). At corresponding interventions, *restriction* maps the sender's measure χ^ρ to a backbone measure χ^τ; *extension* maps χ^τ to a recipient measure $\chi^{\rho,\sigma}$ that restricts back to χ^τ (D'Acunto & Battiloro, 2025, Defs. 8–9, Eq. 9). We write $\chi^\tau \leftrightarrow Q$ when collecting backbone measures into a response kernel. Appendix A provides further background on RCK.

4 BACKBONE IDENTIFIABILITY UNDER PRIVATE KNOWLEDGE

We now formalise a distributed RCK setting in which agents hold private causal knowledge about different causes of a shared outcome. Communication makes a sender's causal perspective available to a recipient, who must interpret it within its own model to perform causal inference on their own interventions of interest. In our formulation, each agent's kernel describes how the shared outcome responds to interventions on its own cause. We ask whether these reports identify the joint-response kernel on the shared backbone.

Formally, Agent ρ reports outcome laws for interventions on Z_ρ and Agent σ for interventions on Z_σ, both with outcome X. The intervention inputs Z_ρ, Z_σ and outcome X take values in arbitrary measurable spaces $\mathcal{Z}_\rho, \mathcal{Z}_\sigma, \mathcal{X}$. They may represent intervention bundles, trajectories, or multivariate outcomes. All events and kernels below refer to the corresponding measurable structures, whose σ-algebras we suppress. Let Q be a candidate kernel for this common backbone, giving the outcome

law under joint interventions:

$$Q : \mathcal{Z}_\rho \times \mathcal{Z}_\sigma \rightsquigarrow \mathcal{X}, \qquad Q(H \mid z_\rho, z_\sigma) = P(X \in H \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)) .$$

For each fixed intervention pair, Q gives the outcome distribution; it does not specify a probability law on the intervention space $\mathcal{Z}_\rho \times \mathcal{Z}_\sigma$. For example, $Q([c, \infty) \mid z_\rho, z_\sigma)$ is the probability of earnings at least c under that joint intervention. The identification target is this backbone response kernel, which retains the relevant outcome laws while omitting internal variables and mechanisms of the local SCMs.

The local report spaces are the families of single-cause intervention kernels

$$\mathcal{K}^\rho := \{K : \mathcal{Z}_\rho \rightsquigarrow \mathcal{X}\}, \qquad \mathcal{K}^\sigma := \{K : \mathcal{Z}_\sigma \rightsquigarrow \mathcal{X}\} .$$

We denote the agents' initially private reports by $\tilde{\chi}^\rho \in \mathcal{K}^\rho, \tilde{\chi}^\sigma \in \mathcal{K}^\sigma$. Each report is assumed to be a valid intervention law, with local confounding addressed by the agent's causal assumptions or identification strategy. Fix background probability laws P_{Z_ρ}, P_{Z_σ}. We represent each agent's local view of a candidate joint response Q by the report operators

$$\pi^\rho(Q)(H \mid z_\rho) := \int Q(H \mid z_\rho, z_\sigma)\, P_{Z_\sigma}(dz_\sigma), \qquad \pi^\sigma(Q)(H \mid z_\sigma) := \int Q(H \mid z_\rho, z_\sigma)\, P_{Z_\rho}(dz_\rho),$$

for each measurable outcome event $H \subseteq \mathcal{X}$. In $\pi^\rho(Q)$, z_ρ remains an input while Z_σ is averaged over; the roles reverse for π^σ. This represents "private" knowledge: each agent initially has access only to its own intervention response family. These fixed averaging relations specify our model of private causes. Each average is a Markov kernel; π^ρ, π^σ are *local report operators*, not RCK restriction maps. For a baseline kernel Q_0, the reports are

$$\tilde{\chi}^\rho := \pi^\rho(Q_0), \qquad \tilde{\chi}^\sigma := \pi^\sigma(Q_0).$$

The agents are not given Q_0.

For identification, both reports and the background laws are taken as known exactly, with specified intervention meanings and outcome scales. No full causal graph, local SCM or joint experimental distribution is supplied. We call a candidate backbone kernel *admissible* if it is a valid Markov kernel (Section 3); it is *compatible* if, in addition, both averages reproduce the supplied reports, for every intervention value and measurable outcome event. The set of all such kernels is

$$\mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma) := \{Q : \mathcal{Z}_\rho \times \mathcal{Z}_\sigma \rightsquigarrow \mathcal{X} : \pi^\rho(Q) = \tilde{\chi}^\rho,\ \pi^\sigma(Q) = \tilde{\chi}^\sigma\}, \tag{2}$$

We use $\mathcal{B}_\tau$ for the same set expressed in RCK edge-value notation. A compatibility set with one member identifies the full kernel. Even when several candidates remain, a particular intervention query is identified if every candidate gives the same answer. Choosing or negotiating one candidate does not itself establish identification; additional information or justified structural restrictions must resolve any ambiguity relevant to the query.

Assumptions. We work under the following conditions:

(A1) there exists at least one admissible kernel $Q_0 \in \mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$;

(A2) there exist measurable sets $A_\rho \subseteq \mathcal{Z}_\rho$ and $A_\sigma \subseteq \mathcal{Z}_\sigma$ with $P_{Z_\rho}(A_\rho) > 0$ and $P_{Z_\sigma}(A_\sigma) > 0$, and bounded measurable functions $u : \mathcal{Z}_\rho \to \mathbb{R}$ and $v : \mathcal{Z}_\sigma \to \mathbb{R}$ satisfying

$$\int u(z_\rho)\, P_{Z_\rho}(dz_\rho) = 0, \qquad \int v(z_\sigma)\, P_{Z_\sigma}(dz_\sigma) = 0,$$

with u not P_{Z_ρ}-almost surely zero, v not P_{Z_σ}-almost surely zero, $u = 0$ outside A_ρ, and $v = 0$ outside A_σ;

(A3) on the rectangle $A_\rho \times A_\sigma$, there exist ν and $\varepsilon > 0$, with ν a probability measure on $\mathcal{X}$, such that $0 < \nu(B) < 1$ for some measurable $B \subseteq \mathcal{X}$ and

$$Q_0(H \mid z_\rho, z_\sigma) \geq \varepsilon\, \nu(H)$$

for every measurable outcome event $H \subseteq \mathcal{X}$ and all $(z_\rho, z_\sigma) \in A_\rho \times A_\sigma$.

Condition (A1) is a baseline compatibility requirement: before asking whether the backbone is unique, the two private reports must admit at least one common backbone candidate.

Condition (A2) is a non-degeneracy assumption selecting a positive-measure rectangle and nonzero, centred functions u, v on the two private-cause spaces. These functions have zero average; the causes need not. A perturbation proportional to $u(z_\rho)v(z_\sigma)$ disappears when either cause is averaged out, preserving both reports, yet can remain nonzero at a fixed joint intervention. Averaging with fixed laws is linear in Q; the causal response need not be linear in its intervention inputs.

Condition (A3) supplies a common, nontrivial component of the outcome laws on $A_\rho \times A_\sigma$. This leaves enough probability mass to perturb while keeping the kernel nonnegative. This bound is a local minorisation (Meyn & Tweedie, 1993, Sec. 5.1); it need not hold over all intervention values. Gaussian or heavy-tailed kernels, and kernels that become near-deterministic elsewhere, are allowed when such a rectangle exists. These are sufficient conditions for the construction, not a necessary-and-sufficient classification of non-identification.

Theorem 1 (Backbone non-identifiability under private knowledge)**.** *Under assumptions* (A1)–(A3)*, there exists a radius $\delta > 0$ and a one-parameter family of distinct kernels,*

$$\{Q_t : t \in (-\delta, \delta)\} \subseteq \mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma),$$

such that:

- *(i) the private reports are identical for every Q_t, that is, $\pi^\rho(Q_t) = \tilde{\chi}^\rho$ and $\pi^\sigma(Q_t) = \tilde{\chi}^\sigma$;*
- *(ii) for $t \neq s$, the kernels Q_t and Q_s induce different joint interventional distributions for some pair (z_ρ, z_σ).*

Since $(-\delta, \delta)$ contains infinitely many distinct parameter values, the same private reports are compatible with infinitely many interventionally distinct backbone candidates.

Proof sketch. Start from $Q_0 \in \mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$ as supplied by (A1). Choose a nonzero signed measure S with $S(\mathcal{X}) = 0$ [Proposition 6]. Using the centred functions u, v from (A2), define

$$Q_t(\cdot \mid z_\rho, z_\sigma) = Q_0(\cdot \mid z_\rho, z_\sigma) + t\, u(z_\rho)v(z_\sigma)S(\cdot). \tag{3}$$

By (A2), $\int u\, dP_{Z_\rho} = 0$ and $\int v\, dP_{Z_\sigma} = 0$, so averaging either input removes the added term and preserves both reports. Since $S(\mathcal{X}) = 0$, each perturbed measure still has total mass one. Assumption (A3) ensures nonnegativity for all sufficiently small $|t| < \delta$, yielding valid kernels [Proposition 7]. Finally, for $t \neq s$, some intervention pair satisfies $Q_t(\cdot \mid z_\rho^\star, z_\sigma^\star) \neq Q_s(\cdot \mid z_\rho^\star, z_\sigma^\star)$. Equivalently, there exists a measurable event $H^\star \subseteq \mathcal{X}$ such that

$$Q_t(H^\star \mid z_\rho^\star, z_\sigma^\star) \neq Q_s(H^\star \mid z_\rho^\star, z_\sigma^\star).$$

Intuitively, the perturbation is invisible to both reports but changes an outcome probability under some joint intervention [Proposition 8]. This proves causal non-identifiability of the backbone kernel Q_0 under assumptions (A1)–(A3). □

All formal proofs are provided in Appendix B.

The witnesses in Appendix C connect the construction to established marginal ambiguities. As in fixed-margin contingency tables (Agresti, 2013), the binary 2×2 reports impose four constraints of rank three, leaving the interaction direction $(1, -1, -1, 1)$ free. In the Gaussian witness, alternating cross-covariances cancel under averaging: covariance summaries remain fixed although the full mixture reports can vary, illustrating why zero correlation does not imply independence (Rényi, 1959). The continuous Laplace witness (§C.3) preserves complete reports beyond finite state spaces and Gaussian response families. Separate interventions can already fail to identify joint effects (Saengkyongam & Silva, 2020). Theorem 1 shows that, on general measurable spaces, every compatible kernel satisfying (A1)–(A3) has arbitrarily close alternatives that preserve both complete local reports but change some joint-intervention probabilities. Restricting the search to nearby RCK backbone candidates therefore does not restore identifiability. A continuum alone follows by mixing two compatible kernels; Appendix E gives a uniform bound and further comparisons. We now link this result back to the RCK language and interpretation.

Proposition 2 (RCK interpretation of Theorem 1)**.** *Let $\mathcal{B}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$ be the compatible backbone values $\chi^\tau \in CK(M^\tau)$ corresponding to kernels in $\mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$. Under assumptions (A1)–(A3),*

$$|\mathcal{B}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)| > 1.$$

Therefore the private reports do not identify a unique common backbone value.

Proposition 2 locates the difficulty in the information the agents exchange. Even after private mechanisms have been abstracted away and valid causal reports obtained, the shared response can remain unidentified. Interventional consistency governs translation but does not establish identifiability of the response from these reports. In Example 2, a teacher effect averaged across neighbourhoods is insufficient to justify its use at a particular disadvantaged exposure. Theorem 1 and Proposition 2 establish this issue for complete distributions, beyond the example's mean-level witness and functional form. The requirement is agreement on the recipient's query; the agents need not disclose or identify their full models.

For a general kernel, Eq. (3) makes the query-specific issue concrete. At a fixed intervention pair and outcome event H, the perturbation changes the requested probability by $t\,u(z_\rho)v(z_\sigma)S(H)$. A nonzero coefficient certifies that the reports leave this answer ambiguous. A zero coefficient means only that this particular perturbation leaves the answer unchanged; identification requires agreement across the whole compatible set. An identified probability from a joint experiment would supply the missing constraint for that query, without necessarily identifying the full kernel. The obstruction concerns this common-effect configuration and its reports. In an unconfounded chain $Z_\rho \to Z_\sigma \to X$, identified causal kernels for the two links instead determine the composed response and the joint-intervention kernel. Appendix E gives the short argument and its information requirements.

Section 5 develops a complementary route: a justified separability assumption, together with suitable causal response information, can make the reduced exchange sufficient for the recipient's calculation.

5 Backbone Identification under Separability and Causal Communication

Theorem 1 shows that identical private reports can fit different joint-intervention kernels. We now restrict candidate backbone kernels to those generated by the additive equation below. Step 1 derives the resulting interventional expectation. Step 2 explains which causal response functions and disturbance law fix the kernel, and why residual summaries do not. Separability is a substantive sufficient assumption, not a rule for selecting the "simplest" compatible model, and the component functions may be nonlinear. It must be justified by domain knowledge or assessed with joint data. It is also not the only possible recovery condition: any justified restriction that leaves a single compatible kernel suffices; for a specified query, agreement of the remaining candidates is enough. For the recovery route in Step 2, the disturbance law must also be fixed, identified, or communicated.

Step 1: additivity of the expected outcome. Here X is real-valued and $f_\rho : \mathcal{Z}_\rho \to \mathbb{R}$, $f_\sigma : \mathcal{Z}_\sigma \to \mathbb{R}$ are measurable. Expectations require integrability under the relevant laws; variance statements additionally require finite second moments.

Proposition 3 (Additive separability yields an additive interventional expectation)**.** *Suppose the shared outcome admits an additively separable structural equation of the form*

$$X = f_\rho(Z_\rho) + f_\sigma(Z_\sigma) + \varepsilon, \tag{4}$$

where ε is an exogenous disturbance with $\mathbb{E}[|\varepsilon|] < \infty$. Then, for every intervention pair $(z_\rho, z_\sigma) \in \mathcal{Z}_\rho \times \mathcal{Z}_\sigma$, the interventional expectation decomposes as

$$\mathbb{E}[X \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)] = f_\rho(z_\rho) + f_\sigma(z_\sigma) + \mathbb{E}[\varepsilon]. \tag{5}$$

The expected outcome in Eq. (5) *has no mixed term in (z_ρ, z_σ). This restricts the expectation; it does not by itself establish uniqueness of the outcome distribution.*

In Eq. (3), the perturbation vanishes from both reports but can remain at a joint intervention. When the relevant first moments exist, it changes the expected outcome by $t\,u(z_\rho)v(z_\sigma)\int x\,S(dx)$. An additive

expectation excludes this mixed contribution when $\int x\, S(dx) \neq 0$, but a distributional perturbation can also leave the expectation unchanged. Proposition 3 therefore restricts the expectation without recovering the full kernel. Step 2 specifies sufficient information to identify the backbone kernel under this separability assumption. Joint experiments could instead test interaction directly, changing the available information.

Example 3 (Education value-added under separability)**.** Continuing Example 2, suppose a unit teacher-quality increase raises expected earnings by α at every neighbourhood exposure, and a unit neighbourhood improvement raises them by β at every teacher exposure. Then $\gamma = 0$: implementing both changes raises expected earnings by $\alpha + \beta$. The education literature motivates considering such additive response models (Hanushek, 1979; Rockoff, 2004; Rivkin et al., 2005). Classroom and neighbourhood mechanisms provide distinct intervention channels, but that distinction does not establish separability. The assumption remains substantive and empirically revisable (Jackson, 2014); the illustration does not test it.

Under separability, changing Z_ρ from z_ρ to z'_ρ changes the expected outcome by $f_\rho(z'_\rho) - f_\rho(z_\rho)$, regardless of z_σ. Each response function may be nonlinear and exhibit diminishing returns. Step 2 asks which communicated quantities suffice to identify the full backbone kernel.

Step 2: recover the kernel from response functions and the disturbance law. Additive separability specifies how the causes' effects combine; it does not tell either agent the other's response function or determine the disturbance law. Subtracting each agent's own response from the outcome gives the residuals

$$R_\rho := X - f_\rho(Z_\rho) = f_\sigma(Z_\sigma) + \varepsilon, \qquad R_\sigma := X - f_\sigma(Z_\sigma) = f_\rho(Z_\rho) + \varepsilon.$$

Each residual still contains the other cause's contribution together with ε. When $Z_\rho \perp\!\!\!\perp Z_\sigma$ and $\varepsilon \perp\!\!\!\perp (Z_\rho, Z_\sigma)$, the residual variances are

$$\mathrm{Var}(R_\rho) = \mathrm{Var}\big(f_\sigma(Z_\sigma)\big) + \mathrm{Var}(\varepsilon), \qquad \mathrm{Var}(R_\sigma) = \mathrm{Var}\big(f_\rho(Z_\rho)\big) + \mathrm{Var}(\varepsilon).$$

The same residual summaries can therefore be compatible with different response functions, and hence with different additive backbone kernels. Under separability, the backbone kernel Q from Section 4 is the disturbance law P_ε shifted by

$$m(z_\rho, z_\sigma) := f_\rho(z_\rho) + f_\sigma(z_\sigma). \tag{6}$$

The expectation under Q is $m(z_\rho, z_\sigma) + \mathbb{E}[\varepsilon]$; differences in m give changes in expected outcomes because $\mathbb{E}[\varepsilon]$ cancels. Explicitly,

$$Q(H \mid z_\rho, z_\sigma) := P_\varepsilon(\{e : m(z_\rho, z_\sigma) + e \in H\}) = P(X \in H \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)), \tag{7}$$

for each measurable outcome event $H \subseteq \mathcal{X}$. Once the recipient has identified f_ρ and f_σ, the shift m is known. The full kernel is determined when the disturbance law is also specified, identified or communicated.

Proposition 4 (Observational summaries remain insufficient under separability)**.** *Assume the additive-separable setting of Proposition 3, and let m be the backbone kernel's deterministic shift in Eq.* (6)*. Then communicating only observational residual summaries such as $\mathrm{Var}(R_\rho)$ and $\mathrm{Var}(R_\sigma)$ does not identify m in general.*

Example 4 (Residual summaries in the education setting)**.** Continuing Example 3 under additive separability, take the linear empirical specification $f_\rho(z_\rho) = \alpha z_\rho$ and $f_\sigma(z_\sigma) = \beta z_\sigma$ with one-year exposures measured in standard-deviation units, so $\mathrm{Var}(Z_\rho) = \mathrm{Var}(Z_\sigma) = 1$. This is already the favourable case in which each causal response is determined by a single slope (see Appendix D). Under $Z_\rho \perp\!\!\!\perp Z_\sigma$ and $\varepsilon \perp\!\!\!\perp (Z_\rho, Z_\sigma)$, writing $\nu_\varepsilon^2 := \mathrm{Var}(\varepsilon)$ gives

$$\mathrm{Var}(R_\rho) = \beta^2 + \nu_\varepsilon^2, \qquad \mathrm{Var}(R_\sigma) = \alpha^2 + \nu_\varepsilon^2.$$

In education terms, R_ρ contains neighbourhood variation through $f_\sigma(Z_\sigma)$, and R_σ contains teacher-quality variation through $f_\rho(Z_\rho)$. Therefore a residual variance alone does not say how much of the remaining variation is due to the other causal channel and how much is due to ε.

Remark 5 (Causal response communication)**.** Agent ρ sends its response kernel $\tilde{\chi}^\rho$ to Agent σ, which retains its own report $\tilde{\chi}^\sigma$ locally. Let $a(z_\rho)$ and $b(z_\sigma)$ be the expected outcomes under the received and local kernels, with finite first moments, and $\mu = \int a\, dP_{Z_\rho} = \int b\, dP_{Z_\sigma}$. Under separability, the recipient identifies the backbone's joint-intervention expectation as $a(z_\rho) + b(z_\sigma) - \mu$ (Appendix D). If the disturbance law is also specified, identified or communicated, the recipient obtains $m(z_\rho, z_\sigma) = a(z_\rho) + b(z_\sigma) - \mu - \mathbb{E}[\varepsilon]$ and hence the full backbone kernel in Eq. (7).

Residual summaries provide strictly less information. The origin of the response functions is not relevant for the identification question. They may be identified from experimental or quasi-experimental designs (Angrist & Pischke, 2009, Chs. 2, 4), and then abstracted to local reports.

Example 5 (Policy-making from a received teacher kernel)**.** Return to Examples 1–2, under the linear separability assumption of Example 4. The teacher researcher sends its response kernel to the neighbourhood researcher, who retains its own kernel locally. With cumulative exposures in standard-deviation years, their expectations a, b from Remark 5 give the contrasts

$$a(z_\rho) - a(0) = 1.3\, z_\rho, \qquad b(z_\sigma) - b(0) = 0.5\, z_\sigma,$$

with effects on expected earnings in percentage points (Chetty et al., 2014b; Chetty & Hendren, 2018). Assume effects add across exposure periods. From its own kernel, the recipient computes a shortfall of $\Delta_\sigma := 18 \times 0.5 = 9$ percentage points for eighteen years at neighbourhood exposure one standard deviation below baseline. Separability makes the received teacher contrast valid at this disadvantaged exposure. For baseline teacher exposure $\bar{z}_\rho$ and compensating exposure z^*_ρ, the recipient solves

$$1.3\,(z^*_\rho - \bar{z}_\rho) = a(z^*_\rho) - a(\bar{z}_\rho) = \Delta_\sigma,$$

so the offsetting teacher input is $\Delta_\sigma/1.3 = \frac{9}{1.3} \approx 6.9$ teacher-quality standard-deviation years. If spread evenly over a 13-year school career, the annual increment is $9/(1.3 \times 13) \approx 0.53$ teacher-quality standard deviations above baseline. Thus the recipient combines its own disadvantage estimate with the received teacher response to identify the compensating intervention left unresolved in Example 2. Under $Z_\rho \perp\!\!\!\perp Z_\sigma$ and $\varepsilon \perp\!\!\!\perp (Z_\rho, Z_\sigma)$, and with observed $\mathrm{Var}(X)$, these linear response contrasts also identify $\mathrm{Var}(\varepsilon)$ (Appendix D).

The examples suggest matching the message to the recipient's question. The compensation in Example 5 uses changes in expected earnings; a policy targeting the probability of earnings below a threshold c instead requires $Q((-\infty, c) \mid z_\rho, z_\sigma)$. Under Eq. (7), this is $P_\varepsilon((-\infty, c - m(z_\rho, z_\sigma)))$, which the disturbance's expectation and variance alone do not specify. The recipient can request this disturbance probability while the detailed local models remain private.

The practical gain in Example 5 is that the neighbourhood researcher identifies a compensating teacher dose from the received kernel, while the classroom model stays local. The other introductory settings can use different recovery structures. Under the chain assumptions in Appendix E, with aligned and identified link kernels, a labour-market model can combine the transmitted test-score kernel with its earnings response, a clinical model can combine a biomarker kernel with its clinical response, and a health model can combine a pollutant kernel with its health response. Those cases use causal composition; our common-outcome recovery uses separability. Both show how identified responses can support a recipient's intervention question without requiring the full local models.

6 Conclusion and Future Work

RCK organises causal communication between complex local models along paths of interventionally consistent translations. Representing selected responses as kernels lets a recipient use them while the sender retains its mechanisms, records and confounding analysis. Our result exposes an identification problem within this reduced exchange: for two causes of a common outcome, every compatible baseline satisfying the stated non-degeneracy and overlap conditions admits arbitrarily close alternatives that preserve both complete reports but disagree on joint interventions.

Additive separability provides one sufficient recovery route: identified response functions fix the joint response surface, and a fixed, identified or communicated disturbance law fixes the full kernel. Residual variances alone are insufficient. Full-kernel ambiguity need not make every policy query ambiguous; what matters for a task is whether all compatible kernels give the same answer.

These results take a step towards operational RCK: determining which causal questions a reduced abstract exchange can answer. Connecting RCK with causal identification and distributed inference could guide what agents need to share while retaining their detailed models locally. Query diagnostics, requests for causal responses or joint evidence, learning abstraction maps, testing separability and extensions to larger networks remain important future directions. The aim is to realise the potential for efficient causal communication with limited disclosure of models and data.

AI USE STATEMENT

Generative AI tools were used to assist with editing and clarifying the exposition and with adapting the manuscript's formatting.

REFERENCES

Alan Agresti. *Categorical Data Analysis*. Wiley, 3 edition, 2013.

Dalal Alrajeh, Hana Chockler, and Joseph Y. Halpern. Combining experts' causal judgments. In *Proc. of AAAI*, pp. 6311–6318, 2018.

Joshua D. Angrist and Jörn-Steffen Pischke. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 2009.

Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proc. of the National Academy of Sciences*, 113(27):7345–7352, 2016. doi: 10.1073/pnas.1510507113.

Sander Beckers and Joseph Y. Halpern. Abstracting causal models. In *Proc. of AAAI*, pp. 2678 – 2685, 2019.

Dimitri P. Bertsekas and Steven E. Shreve. *Stochastic Optimal Control: The Discrete-Time Case*, volume 139 of *Mathematics in Science and Engineering*. Academic Press, 1978.

Richard Bradley, Franz Dietrich, and Christian List. Aggregating causal judgments. *Philosophy of Science*, 81(4):491–515, 2014. doi: 10.1086/678044.

Raj Chetty and Nathaniel Hendren. The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228, 2018. doi: 10.1093/qje/qjy006.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates. *American Economic Review*, 104(9):2593–2632, 2014a. doi: 10.1257/aer.104.9.2593.

Raj Chetty, John N. Friedman, and Jonah E. Rockoff. Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood. *American Economic Review*, 104(9):2633–2679, 2014b. doi: 10.1257/aer.104.9.2633.

Raj Chetty, Nathaniel Hendren, and Lawrence F. Katz. The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment. *American Economic Review*, 106(4):855–902, 2016. doi: 10.1257/aer.20150572.

Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994. doi: 10.1016/0165-1684(94)90029-9.

Veronica Czitrom. One-factor-at-a-time versus designed experiments. *The American Statistician*, 53 (2):126–131, 1999. doi: 10.1080/00031305.1999.10474445.

Gabriele D'Acunto and Claudio Battiloro. The Relativity of Causal Knowledge. In *Proc. of UAI*, pp. 863–881, 2025. URL https://proceedings.mlr.press/v286/d-acunto25a.html.

Gabriele D'Acunto, Fabio Massimo Zennaro, Yorgos Felekis, and Paolo Di Lorenzo. Causal Abstraction Learning based on the Semantic Embedding Principle. In *Proc. of ICML*, 2025.

Gabriele D'Acunto, Paolo Di Lorenzo, and Sergio Barbarossa. Networks of causal abstractions: A sheaf-theoretic framework. *arXiv preprint arXiv:2509.25236*, 2026. URL https://arxiv.org/abs/2509.25236v3. Version 3, revised 28 April 2026.

Department for Education. Apply to access the Longitudinal Education Outcomes (LEO) dataset. Online guidance, 2026a. Updated 14 May 2026. Accessed 25 September 2026.

Department for Education. Pupil Premium: Overview, 2026b. URL https://www.gov.uk/government/publications/pupil-premium/pupil-premium. Updated 24 August 2026. Accessed 25 September 2026.

Frederick Eberhardt, Patrik O. Hoyer, and Richard Scheines. Combining experiments to discover linear cyclic models with latent variables. In *Proc. of AISTATS*, pp. 185–192, 2010. URL https://proceedings.mlr.press/v9/eberhardt10a.html.

Meir Friedenberg and Joseph Y. Halpern. Combining the Causal Judgments of Experts with Possibly Different Focus Areas. In *Proc. of KR*, 2018.

Matthew Friend, Aleks Kissinger, Robert W. Spekkens, and Elie Wolfe. Compositional causal identification from imperfect or disturbing observations. *Entropy*, 27(7):732, 2025.

Sander Greenland. Interactions in Epidemiology: Relevance, Identification, and Estimation. *Epidemiology*, 20(1):14–17, 2009. doi: 10.1097/EDE.0b013e318193e7b5.

Luigi Gresele, Julius von Kügelgen, Jonas Kübler, Elke Kirschbaum, Bernhard Schölkopf, and Dominik Janzing. Causal inference through the structural causal marginal problem. In *Proc. of ICML*, pp. 7793–7824, 2022. URL https://proceedings.mlr.press/v162/gresele22a.html.

Eric A. Hanushek. Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *Journal of Human Resources*, 14(3):351–388, 1979. doi: 10.2307/145575.

Patrik O. Hoyer, Dominik Janzing, Joris M. Mooij, Jonas Peters, and Bernhard Schölkopf. Nonlinear causal discovery with additive noise models. In *Proc. of NeurIPS*, pp. 689–696, 2008.

C. Kirabo Jackson. Teacher Quality at the High-School Level: The Importance of Accounting for Tracks. *Journal of Labor Economics*, 32(4):645–684, 2014. doi: 10.1086/676017.

Yonghan Jung, Jin Tian, and Elias Bareinboim. Estimating Joint Treatment Effects by Combining Multiple Experiments. In *Proc. of ICML*, pp. 15451–15527, 2023. URL https://proceedings.mlr.press/v202/jung23c.html.

Olav Kallenberg. *Foundations of Modern Probability*. Probability and Its Applications. Springer, 2 edition, 2002. doi: 10.1007/978-1-4757-4015-8.

Mirjam J. Knol and Tyler J. VanderWeele. Recommendations for presenting analyses of effect modification and interaction. *International Journal of Epidemiology*, 41(2):514–520, 2012. doi: 10.1093/ije/dyr218.

Sanghack Lee, Juan D. Correa, and Elias Bareinboim. General Identifiability with Arbitrary Surrogate Experiments. In *Proc. of UAI*, pp. 389–398, 2020. URL https://proceedings.mlr.press/v115/lee20b.html.

Loka Li, Ignavier Ng, Gongxu Luo, Biwei Huang, Guangyi Chen, Tongliang Liu, Bin Gu, and Kun Zhang. Federated causal discovery from heterogeneous data. In *Proc. of ICLR*, 2024. URL https://openreview.net/forum?id=m7tJxajC3G.

Riccardo Massidda, Atticus Geiger, Thomas Icard, and Davide Bacciu. Causal abstraction with soft interventions. In *Proc. of CLeaR*, pp. 68–87, 2023. URL https://proceedings.mlr.press/v213/massidda23a.html.

Conor Mayo-Wilson. Causal identifiability and piecemeal experimentation. *Synthese*, 196(8):3029–3065, 2019. doi: 10.1007/s11229-018-1826-4.

Sean P. Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, London, 1993. doi: 10.1007/978-1-4471-3267-7. URL https://probability.ca/MT/.

Roger B. Nelsen. *An Introduction to Copulas*. Springer, 2 edition, 2006. doi: 10.1007/0-387-28678-0.

Office for National Statistics. Instructions and guidance for output, ingest and transfer requests. Secure Research Service guidance, 2025. Version 2, revised 15 December 2025.

Giambattista Parascandolo, Niki Kilbertus, Mateo Rojas-Carulla, and Bernhard Schölkopf. Learning independent causal mechanisms. In *Proc. of ICML*, 2018.

Judea Pearl. *Causality*. Cambridge University Press, 2009.

Judea Pearl and Elias Bareinboim. External validity: From do-calculus to transportability across populations. *Statistical Science*, 29(4):579–595, 2014. doi: 10.1214/14-STS486.

Jonas Peters, Joris M. Mooij, Dominik Janzing, and Bernhard Schölkopf. Identifiability of causal graphs using functional models. In *Proc. of UAI*, pp. 589–598, 2011.

Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 2017.

Alfréd Rényi. On measures of dependence. *Acta Mathematica Academiae Scientiarum Hungaricae*, 10(3–4):441–451, 1959. doi: 10.1007/BF02024507.

Ella F. Rischel and Sebastian Weichwald. Compositional abstraction error and a category of causal models. In *Proc. of UAI*, pp. 1013–1023, 2021.

Steven G. Rivkin, Eric A. Hanushek, and John F. Kain. Teachers, schools, and academic achievement. *Econometrica*, 73(2):417–458, 2005. doi: 10.1111/j.1468-0262.2005.00584.x.

Jonah E. Rockoff. The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, 94(2):247–252, 2004. doi: 10.1257/0002828041302244.

Mateo Rojas-Carulla, Bernhard Schölkopf, Richard Turner, and Jonas Peters. Invariant models for causal transfer learning. *Journal of Machine Learning Research*, 19(36):1–34, 2018.

Sorawit Saengkyongam and Ricardo Silva. Learning Joint Nonlinear Effects from Single-variable Interventions in the Presence of Hidden Confounders. In *Proc. of UAI*, pp. 300–309, 2020. URL `https://proceedings.mlr.press/v124/saengkyongam20a.html`.

Bernhard Schölkopf, Dominik Janzing, Jonas Peters, Eleni Sgouritsa, Kun Zhang, and Joris M. Mooij. On causal and anticausal learning. In *Proc. of ICML*, 2012.

Shohei Shimizu, Patrik O. Hoyer, Aapo Hyvärinen, and Antti Kerminen. A Linear Non-Gaussian Acyclic Model for Causal Discovery. *Journal of Machine Learning Research*, 7(72):2003–2030, 2006.

Jin Tian and Judea Pearl. On the Testable Implications of Causal Models with Hidden Variables. In *Proc. of UAI*, pp. 519–527, 2002.

C. F. Jeff Wu and Michael S. Hamada. *Experiments: Planning, Analysis, and Optimization*. Wiley, 2 edition, 2009. ISBN 9780471699460.

Ruoxuan Xiong, Allison Koenecke, Michael Powell, Zhu Shen, Joshua T. Vogelstein, and Susan Athey. Federated Causal Inference in Heterogeneous Observational Data. *Statistics in Medicine*, 42(24):4418–4439, 2023. doi: 10.1002/sim.9868.

(a) Before communication: different intervention knowledge

The agents have local SCMs and a common outcome X, but each initially lacks the other's intervention responses.

Agent ρ	**Agent** σ
Knows how X responds to $\mathrm{do}(Z_\rho = z_\rho)$.	Knows how X responds to $\mathrm{do}(Z_\sigma = z_\sigma)$.
Lacks the responses to $\mathrm{do}(Z_\sigma = z_\sigma)$.	Lacks the responses to $\mathrm{do}(Z_\rho = z_\rho)$.

Agent ρ sends its response kernel to Agent σ, which retains its own report locally. The recipient asks how X responds when both causes are set; full SCMs and records remain local.

(b) RCK: translating an available causal response

With a supplied shared abstraction and maps, Agent ρ sends a reduced causal value for Agent σ to interpret in its own model:

$$\underbrace{\chi^{\rho}}_{\text{sender's measure}} \xrightarrow{\text{restrict}} \underbrace{\chi^{\rho,\tau}}_{\text{abstract message}} \xrightarrow{\text{extend}} \underbrace{\chi^{\rho,\sigma}}_{\text{recipient's interpretation}} .$$

These are measures at corresponding interventions; collecting the selected outcome laws across interventions gives kernels. This transfer communicates available responses. It does not, by itself, identify a new joint response.

(c) After receipt: what can Agent σ identify?

The recipient has the received kernel $\tilde{\chi}^{\rho}$ and its own $\tilde{\chi}^{\sigma}$, on specified intervention and outcome scales. Each report averages over the other cause with a fixed known law. A candidate joint response

$$Q(H \mid z_\rho, z_\sigma) = P(X \in H \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma))$$

must reproduce both reports:

$$\tilde{\chi}^{\rho}(H \mid z_\rho) = \int Q(H \mid z_\rho, z_\sigma)\, P_{Z_\sigma}(dz_\sigma),$$

$$\tilde{\chi}^{\sigma}(H \mid z_\sigma) = \int Q(H \mid z_\rho, z_\sigma)\, P_{Z_\rho}(dz_\rho).$$

Under (A1)–(A3), arbitrarily close kernels reproduce both reports but disagree on joint interventions. Receiving the sender's kernel can therefore leave the recipient's joint-intervention query unresolved.

Figure 1: From local causal knowledge to communication and joint-response identification. Panels (a) and (c) specify our common-outcome scenario; panel (b) gives RCK's transfer rule (D'Acunto & Battiloro, 2025, Sec. 4). RCK itself does not impose the two-report restriction or the averaging equations in (c). Locality concerns access to intervention responses; intervention meanings are assumed specified. Report averaging is distinct from RCK restriction.

A RCK BACKGROUND

Figure 1 connects the need to communicate locally held responses with RCK transport and the identification question that remains after exchange. This appendix connects the kernel representation to RCK's original definitions and records the maps used for communication.

A.1 FROM LOCAL SCMS TO INTERVENTIONAL KERNELS

RCK's Definition 1, *Measure-theoretic SCM*, specifies exogenous and endogenous probability spaces and measurable functions expressing endogenous values in terms of exogenous ones (D'Acunto & Battiloro, 2025). It assumes independent exogenous variables, and the paper develops its construction for continuous variables. Our general SCM notation allows a joint exogenous law; the

non-identification theorem works directly with response kernels under (A1)–(A3). The following construction explains how an SCM supplies those responses. Let $\mathcal{V}^\rho$ denote the complete endogenous state space of a local SCM, and let U^ρ have law P_{U^ρ}. For an intervention i in a measurable intervention space $\mathcal{I}^\rho$, write g_i^ρ for the measurable solution map obtained from its intervened structural equations. Its full interventional law is the pushforward

$$\mu_i^\rho = (g_i^\rho)_{\#} P_{U^\rho}, \qquad \mu_i^\rho(A) = P_{U^\rho}\big((g_i^\rho)^{-1}(A)\big).$$

For a measurable outcome map $h_\rho : \mathcal{V}^\rho \to \mathcal{X}$, the response on the chosen outcome space is

$$K_\rho(H \mid i) = \mu_i^\rho\big(h_\rho^{-1}(H)\big).$$

If $i \mapsto K_\rho(H \mid i)$ is measurable for each measurable H, this family is a Markov kernel. Local mediators, common causes and other variables are integrated out in this outcome law, while their causal influence remains encoded in it. Computing or identifying the law still requires a valid local model or identification argument. Reducing an observational conditional by the same outcome map does not turn it into an intervention law.

In the original RCK formulation, $CK(M)$ first denotes the category containing an SCM and its intervened states, with interventions as arrows (D'Acunto & Battiloro, 2025, Def. 4). RCK then represents models by convex spaces of probability measures and interventionally consistent abstractions by affine (mixture-preserving) measurable maps (D'Acunto & Battiloro, 2025, Sec. 3, Theorem 3). RCK also uses $CK(M)$ for this distributional representation in its network construction (D'Acunto & Battiloro, 2025, Sec. 4, App. E); we follow that usage. The notation χ^τ for an individual backbone measure is likewise inherited from RCK. At a fixed intervention, the response kernel gives the corresponding outcome law, as in the construction above. The shorthand $\chi^\tau \leftrightarrow Q$ collects these laws as the intervention input varies; it does not identify a single measure with a kernel. The RCK maps below act at corresponding interventions. A response kernel preserves these outcome laws without specifying the local structural equations or a coupling of counterfactual outcomes across interventions.

A.2 Interventional consistency and reduction

Let $\mathcal{V}^\tau$ be the endogenous state space of the shared model M^τ, let $a_\rho : \mathcal{V}^\rho \to \mathcal{V}^\tau$ be a measurable state-abstraction map and let ω_ρ associate the local interventions being represented with their abstract counterparts. At the level of interventional distributions, consistency requires

$$(a_\rho)_{\#}\mu_i^\rho = \mu^\tau_{\omega_\rho(i)} \qquad \text{for every represented intervention } i. \tag{8}$$

Thus, intervene and then abstract gives the same law as evaluating the corresponding intervention in the abstract model. In particular, two represented local interventions mapped to the same abstract intervention must have the same abstracted outcome law. RCK's full abstraction definition also includes maps for the exogenous variables and requires consistency with interventions (D'Acunto & Battiloro, 2025, Defs. 5–6); Eq. (8) states the distributional requirement used here.

Projecting $\mu^\tau_{\omega_\rho(i)}$ onto the selected outcome gives a response kernel such as Q. This is a reduction in the causal information that must be shared, not a claim of a uniquely minimal abstraction or a finite-dimensional message. More than one local SCM can induce the same selected responses. The recipient may use these responses while retaining its own remaining mechanisms; the sender's confounding assumptions are still needed to justify the responses in the first place.

A.3 Restriction, extension and network alignment

In a network $N = (\mathcal{N}, \mathcal{E})$, node ρ carries a local SCM M^ρ, and edge $\tau : \rho \sim \sigma$ carries an abstract SCM M^τ, related to both local SCMs by interventionally consistent abstraction maps. For a node ρ connected to an edge τ, written $\rho \triangleleft \tau$, RCK specifies a restriction map $\alpha_{\mathcal{X}}^{\rho\triangleleft\tau} : CK(M^\rho) \to CK(M^\tau)$. In the figure, this is the step that takes Agent ρ's local causal value χ^ρ and projects the part relevant to the shared outcome space $\mathcal{X}$ into the backbone τ. Information is interpreted in the receiving model through extension maps $\beta_{\mathcal{X}}^{\sigma\triangleleft\tau} : CK(M^\tau) \to CK(M^\sigma)$, shown by the arrow that transports backbone-level content into Agent σ's perspective.

The maps are specified when constructing the shared abstraction; they are not determined by the network incidence relation alone. The subscript $\mathcal{X}$ denotes the endogenous component of the

interventionally consistent abstraction map (Rischel & Weichwald, 2021; D'Acunto & Battiloro, 2025). The image of the restriction–extension composite is the transported knowledge space $CK^{\rho,\sigma} \subseteq CK(M^\sigma)$; an individual transported value is denoted $\chi^{\rho,\sigma}$.

For a specified edge value χ^τ, an extension must return a recipient value consistent with it:

$$\beta_{\mathcal{X}}^{\sigma\triangleleft\tau}(\chi^\tau) \in \{\eta \in CK(M^\sigma) : \alpha_{\mathcal{X}}^{\sigma\triangleleft\tau}(\eta) = \chi^\tau\}.$$

This is the recipient-side compatibility condition in D'Acunto & Battiloro (2025, Eq. 9). It does not assert that restriction has a unique inverse: the extension uses the recipient's model, and the existence and choice of suitable maps are part of the RCK construction. Agreement of the agents' current values means that they induce the same measure after restriction to the edge:

$$\alpha_{\mathcal{X}}^{\rho\triangleleft\tau}(\chi^\rho) = \alpha_{\mathcal{X}}^{\sigma\triangleleft\tau}(\chi^\sigma) = \chi^\tau.$$

This is RCK's agreement condition for the current local values (D'Acunto & Battiloro, 2025, Def. 10). The alignment needed for communication is provided by the shared abstraction and maps relating the agents' causal representations (D'Acunto & Battiloro, 2025, App. E). These maps take the sender's measure to its interpretation in the recipient's model (D'Acunto & Battiloro, 2025, Def. 11):

$$\chi^{\rho,\sigma} = \beta_{\mathcal{X}}^{\sigma\triangleleft\tau}\big(\alpha_{\mathcal{X}}^{\rho\triangleleft\tau}(\chi^\rho)\big).$$

The transported value $\chi^{\rho,\sigma}$ restricts to the sender's backbone message; the recipient's current χ^σ need not already project to that same message. Formally, the incidence order on $\Pi = \mathcal{N} \cup \mathcal{E}$ has $\rho \triangleleft \tau$ when node ρ is incident to edge τ. The RCK sheaf assigns a convex space of probability measures to each node and edge, and an affine restriction map to each node–edge incidence. Its cosheaf reverses the map direction, using extensions whose images embed in the recipient's space (D'Acunto & Battiloro, 2025, Defs. 8–9). A *global section* is a collection of node values satisfying the displayed agreement condition on every edge (D'Acunto & Battiloro, 2025, Def. 10). This states simultaneous consistency across the network.

For a path $\rho_0, \tau_1, \rho_1, \ldots, \tau_k, \rho_k$, define

$$\mathcal{T}_j := \beta_{\mathcal{X}}^{\rho_j\triangleleft\tau_j} \circ \alpha_{\mathcal{X}}^{\rho_{j-1}\triangleleft\tau_j}. \qquad \chi^{\rho_0,\rho_k} = (\mathcal{T}_k \circ \cdots \circ \mathcal{T}_1)(\chi^{\rho_0}).$$

This is RCK transport along a path (D'Acunto & Battiloro, 2025, Def. 11); different paths can produce different recipient interpretations. These constructions organise supplied causal values and maps across a network. In our scenario, communication makes another agent's previously unavailable intervention responses accessible. The remaining question is whether those responses, together with the recipient's own, determine the joint response needed for the task. RCK specifies transport given shared abstractions and maps; it does not stipulate that each agent observes exactly one of our averaged report families. That restriction is our information model. We do not model ignorance of the intervention meanings themselves, or derive this information restriction from a particular pair of full SCMs. Our proof concerns kernels satisfying the two report equations. If suitable transport maps are available for competing candidates, successful transfer alone does not select between their different joint responses. Additional information encoded by a particular local SCM or supplied map must be treated as an additional constraint, beyond the reports analysed here.

Compatibility in RCK notation. Writing the same family of responses in RCK notation, the counterpart of Eq. (2) is

$$\mathcal{B}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma) := \{\chi^\tau \in CK(M^\tau) : \pi^\rho(\chi^\tau) = \tilde{\chi}^\rho,\ \pi^\sigma(\chi^\tau) = \tilde{\chi}^\sigma\}. \tag{9}$$

Under $\chi^\tau \leftrightarrow Q$, this is the same identification question in RCK notation. It imposes the two report constraints, not access to the full local SCMs. Compatibility alone neither constructs abstraction maps nor guarantees additional restrictions a particular application might impose. In particular, the fixed averaging laws in Section 4 specify the information retained in the reports. If an intervention changes the other channel's distribution, its induced averaging law must be modelled explicitly. A particular pair of full local SCMs may also impose constraints beyond these two report equalities; the theorem concerns the stated kernel compatibility class. Restriction and extension are used only once the shared abstraction and suitable maps have been specified. Two compatible kernels can disagree on a query even when transport through each supplied abstraction is well defined.

Worked education interface. Example 1 is motivated by actual controls on administrative-data research. England's Longitudinal Education Outcomes data link education to employment and earnings, with access limited to accredited projects and requested variables (Department for Education, 2026a). Secure research services require outputs and transfers to meet project permissions and disclosure controls (Office for National Statistics, 2025). These arrangements motivate a recipient who can obtain approved findings without access to the underlying study. The example assumes that the teacher's approved message is an interventional earnings kernel, while records and full SCMs remain local. This is a specified information boundary, not a universal ban on model sharing or a privacy guarantee for kernels. We grant valid local causal responses, supplied RCK maps and known intervention meanings; disclosure clearance establishes none of these. As in Section 4, reports are idealised as known exactly, separating identification from estimation and disclosure control. Pupil Premium motivates the question; the US studies do not evaluate or calibrate that English programme.

In Example 1, the teacher researcher's local SCM may include baseline attainment and family background B, teacher assignments and exposures $T_{1:13}$ over thirteen school years, learning and test scores $L_{1:13}$, and an earnings measure Y_ρ. The neighbourhood researcher's SCM may instead include family resources F, move timing A_m, neighbourhood histories $N_{0:18}$, and an earnings measure Y_σ. Their observational and interventional laws involve these full collections. For instance, baseline family conditions may affect both teacher assignment and later earnings. Under an appropriate sufficient-adjustment assumption with covariates C_ρ, a reduced causal kernel can be obtained as

$$K_\rho(H \mid z_\rho) = \int P_\rho(X \in H \mid Z_\rho = z_\rho, C_\rho = c)\, P_{C_\rho}(dc).$$

Here P_ρ is the teacher study's observational law and P_{C_ρ} its covariate marginal. The formula requires consistency (observed outcomes agree with those under the received intervention), conditional exchangeability (no remaining confounding given C_ρ) and positivity (support for the intervention values in each relevant covariate stratum) (Pearl, 2009); another local identification strategy may be needed. Once this intervention law is justified, the recipient can use it without receiving C_ρ, its adjustment model or the records. These adjustment variables differ from the private cause averaged out by π^ρ. The resulting kernel can serve as a report in Section 4 when the stated averaging relations also hold. The teacher study includes a dynamic model of classroom assignment, test scores and earnings (Chetty et al., 2014b, Online Appendix A); the neighbourhood models account for exposure histories, move timing and family circumstances (Chetty et al., 2016; Chetty & Hendren, 2018). Our trajectories illustrate the larger policy setting, rather than reproduce the studies' exact specifications. Even a union of the graphs would not specify common measurement and intervention meanings, a joint earnings mechanism, or a joint law for potentially overlapping unobserved factors.

For the subsequent identification question, we index the translated teacher and neighbourhood interventions by exposures Z_ρ, Z_σ, with the earnings measures mapped to X. In the scalar illustration these exposures might summarise trajectories, but equal cumulative exposure need not imply equal effects: temporal aggregation and the intervention correspondence require substantive justification. Example 1 assumes a shared interventionally consistent abstraction and supplied maps; the identification illustration additionally assumes a harmonised target population. The separate studies do not establish these maps or provide the full report kernels assumed here. In particular, if two local interventions map to the same abstract intervention, their translated outcome distributions must agree. Preserving only an observational distribution or an interventional mean would not ensure consistency of the full response kernel.

After this translation and marginalisation of the remaining local variables, the teacher researcher sends the single-cause kernel $\tilde{\chi}^\rho(dx \mid z_\rho)$ to the neighbourhood researcher, who already holds $\tilde{\chi}^\sigma(dx \mid z_\sigma)$ locally. The recipient's task is to identify the backbone kernel $Q(dx \mid z_\rho, z_\sigma)$, or its required query, using the incoming teacher response and its own neighbourhood response. This kernel specifies the earnings distribution when both exposures are set, allowing the recipient to evaluate a teacher intervention at a disadvantaged neighbourhood exposure. The classroom mechanisms and records stay with the teacher researcher; the neighbourhood report need not be sent back.

Receipt alone does not resolve this identification task: the teacher kernel averages over neighbourhood exposure and need not determine the teacher response at the recipient's chosen exposure. Example 2 illustrates this ambiguity for expectations, and Theorem 1 establishes it for complete reports under (A1)–(A3). Under separability, Remark 5 shows how the recipient identifies the backbone's expected outcome from the received and local kernels. Example 5 then uses the teacher contrast from the

incoming kernel together with the disadvantage quantified by the recipient's own kernel to identify a compensating intervention. If the disturbance law is also known, the same construction recovers the full backbone kernel. The benefit of abstraction is to limit what must be shared; the recipient must still establish that the reduced information identifies its causal query.

In the further examples following Example 1, restriction expresses findings on a common test-score, pollutant or biomarker scale, and extension uses the recipient's compatible downstream causal model. Aligning units or calibrating a sensor is only part of this construction: the abstraction must preserve the relevant intervention responses. These mediator examples motivate communication generally; they need not have the common-outcome structure of the non-identification theorem.

B PROOFS

This appendix gives the full proofs of Theorem 1, and Propositions 2–4 in the main text.

B.1 PROOF OF THEOREM 1

The proof is organised around three tasks. First, we need a perturbation direction on the outcome space that changes some joint response but preserves total mass; that is the role of Proposition 6. Second, we need to lift that direction into a family of admissible backbone kernels that still reproduces the same private reports; that is the role of Proposition 7. Third, we need to show that different members of this family are not merely different parameterisations of the same object, but disagree on at least one joint intervention; that is the role of Proposition 8. For a bounded function h, write $\|h\|_\infty = \sup_x |h(x)|$. A finite signed measure can assign negative mass; $|S|$ denotes its total-variation measure. For $S(H) = \int_H s\, d\nu$, this is $|S|(H) = \int_H |s|\, d\nu$.

Proposition 6 (Signed perturbation on the outcome space)**.** *Let ν be a probability measure on $\mathcal{X}$. If there exists a measurable set $B \subseteq \mathcal{X}$ with $0 < \nu(B) < 1$, then there exists a bounded measurable function $s : \mathcal{X} \to \mathbb{R}$ such that*

(i) $\int s(x)\, \nu(dx) = 0$*;*

(ii) $\|s\|_\infty \le 1$*;*

(iii) $s \not\equiv 0$ *ν-almost surely.*

Consequently, the signed measure

$$S(H) := \int_H s(x)\, \nu(dx) \tag{10}$$

has the following properties:

(a) $S(\mathcal{X}) = 0$*;*

(b) $S \neq 0$*;*

(c) $|S|(H) \le \nu(H)$ *for every measurable set $H \subseteq \mathcal{X}$.*

Proof. Choose a measurable set $B \subseteq \mathcal{X}$ with $0 < \nu(B) < 1$ and define

$$s(x) := \mathbf{1}_B(x) - \nu(B),$$

where $\mathbf{1}_B(x)$ denotes the indicator of B, equal to 1 on B and 0 outside B. We verify the properties in order. For (i),

$$\int s\, d\nu = \int \mathbf{1}_B\, d\nu - \nu(B) \int 1\, d\nu = \nu(B) - \nu(B) \cdot 1 = 0.$$

Property (ii) holds because $\mathbf{1}_B(x) \in \{0, 1\}$, so $s(x)$ is either $1 - \nu(B)$ or $-\nu(B)$, and both have absolute value at most one. Property (iii) holds because $0 < \nu(B) < 1$, so s is positive on B and negative on B^c, both of which have positive ν-measure.

For S as defined in Eq. (10), property (a) is exactly (i) evaluated on $H = \mathcal{X}$. For property (b), consider the measurable set $H_+ := \{x : s(x) > 0\}$. Because s is not zero ν-almost surely (by (iii)), either $\nu(H_+) > 0$ or $\nu(\{x : s(x) < 0\}) > 0$. In the first case, $S(H_+) = \int_{H_+} s\, d\nu > 0$; in the second case, taking $H_- := \{x : s(x) < 0\}$ gives $S(H_-) < 0$. Hence $S \neq 0$.

Finally, for (c),

$$|S|(H) = \int_H |s(x)|\, \nu(dx) \leq \int_H 1\, \nu(dx) = \nu(H),$$

where we use $\|s\|_\infty \leq 1$ from (ii) and $|s(x)| \leq \|s\|_\infty$. □

This proposition is the seed of the perturbation argument. The later kernel construction will add the perturbation S to Q_0 pointwise at each intervention pair (z_ρ, z_σ). For Q_0 to remain a probability kernel, the perturbation must change some events but contribute zero net mass overall; that is exactly what the zero-total-mass property of S just proven in this proposition guarantees. We next look at the perturbation effect on the marginals.

Proposition 7 (Perturbation family with fixed marginals)**.** *The event-level form of Eq.* (3) *is*

$$Q_t(H \mid z_\rho, z_\sigma) := Q_0(H \mid z_\rho, z_\sigma) + t\, u(z_\rho) v(z_\sigma) S(H). \tag{11}$$

Then for every

$$|t| < \delta := \frac{\varepsilon}{\|u\|_\infty \|v\|_\infty},$$

Q_t *is a probability kernel and belongs to the kernel-level compatibility set* $\mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$ *defined in Eq.* (2)*.*

Proof. Fix (z_ρ, z_σ). To show that $Q_t(\cdot \mid z_\rho, z_\sigma)$ is a probability measure, we check normalisation and nonnegativity.

For normalisation, evaluate the definition of Q_t on the whole outcome space:

$$Q_t(\mathcal{X} \mid z_\rho, z_\sigma) = Q_0(\mathcal{X} \mid z_\rho, z_\sigma) + t\, u(z_\rho) v(z_\sigma) S(\mathcal{X}).$$

Because Q_0 is an admissible kernel, evaluating it on the whole outcome space gives $Q_0(\mathcal{X} \mid z_\rho, z_\sigma) = 1$. Because Proposition 6(a) gives $S(\mathcal{X}) = 0$, the perturbation contributes no net mass. Hence

$$Q_t(\mathcal{X} \mid z_\rho, z_\sigma) = 1.$$

For nonnegativity, consider two cases. If $(z_\rho, z_\sigma) \notin A_\rho \times A_\sigma$, then by the support condition in (A2), $u(z_\rho)v(z_\sigma) = 0$, so $Q_t(\cdot \mid z_\rho, z_\sigma) = Q_0(\cdot \mid z_\rho, z_\sigma)$ is a probability measure trivially. If $(z_\rho, z_\sigma) \in A_\rho \times A_\sigma$, the local minorisation condition (A3) applies through its lower-bound inequality. Using the elementary bounds $|u(z_\rho)| \leq \|u\|_\infty$, $|v(z_\sigma)| \leq \|v\|_\infty$, together with the defining total-variation inequality $S(H) \geq -|S|(H)$ for signed measures:

$$Q_t(H \mid z_\rho, z_\sigma) \geq Q_0(H \mid z_\rho, z_\sigma) - |t|\, \|u\|_\infty \|v\|_\infty\, |S|(H),$$

for every measurable set $H \subseteq \mathcal{X}$. By assumption (A3), $Q_0(H \mid z_\rho, z_\sigma) \geq \varepsilon \nu(H)$ on this rectangle, and by Proposition 6(c), $|S|(H) \leq \nu(H)$. Substituting these two bounds gives

$$Q_t(H \mid z_\rho, z_\sigma) \geq \big(\varepsilon - |t|\, \|u\|_\infty \|v\|_\infty\big)\, \nu(H) \geq 0$$

whenever $|t| < \delta = \varepsilon / (\|u\|_\infty \|v\|_\infty)$, because then $\varepsilon - |t|\, \|u\|_\infty \|v\|_\infty > 0$ and $\nu(H) \geq 0$. So $Q_t(\cdot \mid z_\rho, z_\sigma)$ is a probability measure for every (z_ρ, z_σ).

Integrating the perturbed kernel of Eq. (3) over P_{Z_σ} gives

$$\int Q_t(\cdot \mid z_\rho, z_\sigma)\, P_{Z_\sigma}(dz_\sigma) = \int Q_0(\cdot \mid z_\rho, z_\sigma)\, P_{Z_\sigma}(dz_\sigma) + t\, u(z_\rho) \Big(\int v(z_\sigma)\, P_{Z_\sigma}(dz_\sigma) \Big) S.$$

The second term vanishes by the centring condition in (A2), so $\pi^\rho(Q_t) = \pi^\rho(Q_0) = \tilde{\chi}^\rho$. The same calculation integrating over P_{Z_ρ} gives $\pi^\sigma(Q_t) = \pi^\sigma(Q_0) = \tilde{\chi}^\sigma$. Hence $Q_t \in \mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$. □

The construction gives a family of distinct kernels that reproduce both complete reports: averaging either centred factor removes the added term. We next show that these kernels disagree on a joint intervention, so the reports leave a causal response undetermined.

Proposition 8 (Interventional non-equivalence). *For $t \neq s$, there exists a pair $(z_\rho^\star, z_\sigma^\star)$ such that*

$$Q_t(\cdot \mid z_\rho^\star, z_\sigma^\star) \neq Q_s(\cdot \mid z_\rho^\star, z_\sigma^\star),$$

and therefore the two admissible backbones induce different joint interventional distributions under the joint intervention $\mathrm{do}(Z_\rho = z_\rho^\star, Z_\sigma = z_\sigma^\star)$.

Proof. By assumption (A2), u is not P_{Z_ρ}-almost surely zero and vanishes outside A_ρ, while v is not P_{Z_σ}-almost surely zero and vanishes outside A_σ. Therefore there exist $z_\rho^\star \in A_\rho$ and $z_\sigma^\star \in A_\sigma$ such that $u(z_\rho^\star) \neq 0$ and $v(z_\sigma^\star) \neq 0$, and hence

$$u(z_\rho^\star)v(z_\sigma^\star) \neq 0.$$

By Proposition 6(b), the signed measure S is nonzero, so there exists a measurable set $H^\star \subseteq \mathcal{X}$ with $S(H^\star) \neq 0$. Then

$$Q_t(H^\star \mid z_\rho^\star, z_\sigma^\star) - Q_s(H^\star \mid z_\rho^\star, z_\sigma^\star) = (t-s)\, u(z_\rho^\star)v(z_\sigma^\star)S(H^\star) \neq 0.$$

So Q_t and Q_s are distinct kernels, and they differ on the interventional distribution associated with the joint intervention $(z_\rho^\star, z_\sigma^\star)$. □

Given these components, we can now prove the core non-identifiability result of Theorem 1.

Theorem 1 (Backbone non-identifiability under private knowledge). *Under assumptions* (A1)–(A3)*, there exists a radius $\delta > 0$ and a one-parameter family of distinct kernels,*

$$\{Q_t : t \in (-\delta, \delta)\} \subseteq \mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma),$$

such that:

(i) the private reports are identical for every Q_t, that is, $\pi^\rho(Q_t) = \tilde{\chi}^\rho$ and $\pi^\sigma(Q_t) = \tilde{\chi}^\sigma$;

(ii) for $t \neq s$, the kernels Q_t and Q_s induce different joint interventional distributions for some pair (z_ρ, z_σ).

Since $(-\delta, \delta)$ contains infinitely many distinct parameter values, the same private reports are compatible with infinitely many interventionally distinct backbone candidates.

Proof of Theorem 1. Proposition 6 provides a nontrivial signed measure S on the outcome space with total mass zero. This gives a direction in which one can perturb an admissible kernel without breaking the normalisation condition. Proposition 7 shows that, after modulating this direction by the centred functions u and v, one obtains a whole one-parameter family of admissible kernels $\{Q_t : |t| < \delta\} \subseteq \mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$ that all reproduce the same observed private reports. Hence the private reports do not identify a unique backbone candidate. Proposition 8 then shows that distinct members of this family disagree on some joint intervention. Therefore the ambiguity is causally meaningful rather than merely representational. This proves non-identifiability of every baseline kernel satisfying assumptions (A1)–(A3) from its two private reports. □

B.2 Proof of Proposition 2

Proposition 2 (RCK interpretation of Theorem 1). *Let $\mathcal{B}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$ be the compatible backbone values $\chi^\tau \in CK(M^\tau)$ corresponding to kernels in $\mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$. Under assumptions (A1)–(A3),*

$$|\mathcal{B}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)| > 1.$$

Therefore the private reports do not identify a unique common backbone value.

Proof. By Propositions 7 and 8, there exist distinct kernels $Q_t, Q_s \in \mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$ with $t \neq s$ that reproduce the same private reports $\tilde{\chi}^\rho$ and $\tilde{\chi}^\sigma$ and disagree on some joint intervention. In the private-cause representation, these kernels are representatives of distinct edge-level candidates $\chi_t^\tau, \chi_s^\tau \in CK(M^\tau)$. Since membership in $\mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$ means that each kernel projects to $\tilde{\chi}^\rho$ under π^ρ and to $\tilde{\chi}^\sigma$ under π^σ, both edge-level candidates lie in $\mathcal{B}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)$. Because Q_t and Q_s differ on some joint intervention, the corresponding edge-level candidates are distinct. Hence

$$|\mathcal{B}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)| > 1.$$

□

B.3 Proof of Proposition 3

Proposition 3 (Additive separability yields an additive interventional expectation)**.** *Suppose the shared outcome admits an additively separable structural equation of the form*

$$X = f_\rho(Z_\rho) + f_\sigma(Z_\sigma) + \varepsilon, \tag{4}$$

where ε is an exogenous disturbance with $\mathbb{E}[|\varepsilon|] < \infty$. Then, for every intervention pair $(z_\rho, z_\sigma) \in \mathcal{Z}_\rho \times \mathcal{Z}_\sigma$, the interventional expectation decomposes as

$$\mathbb{E}[X \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)] = f_\rho(z_\rho) + f_\sigma(z_\sigma) + \mathbb{E}[\varepsilon]. \tag{5}$$

The expected outcome in Eq. (5) *has no mixed term in (z_ρ, z_σ). This restricts the expectation; it does not by itself establish uniqueness of the outcome distribution.*

Proof. Under the intervention $\mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)$, the structural equation becomes

$$X = f_\rho(z_\rho) + f_\sigma(z_\sigma) + \varepsilon.$$

Taking expectations yields Eq. (5). In Eq. (3), the change in the expected outcome is $t\,u(z_\rho)v(z_\sigma)\int x\,S(dx)$ when the relevant first moments exist. An additive expectation excludes this mixed contribution if $\int x\,S(dx) \neq 0$; a perturbation with zero first moment can still change the outcome law without changing its expectation. □

B.4 Proof of Proposition 4

Proposition 4 (Observational summaries remain insufficient under separability)**.** *Assume the additive-separable setting of Proposition 3, and let m be the backbone kernel's deterministic shift in Eq.* (6)*. Then communicating only observational residual summaries such as $\mathrm{Var}(R_\rho)$ and $\mathrm{Var}(R_\sigma)$ does not identify m in general.*

Proof. It suffices to give a counterexample. Let

$$Z_\rho, Z_\sigma \overset{\text{iid}}{\sim} \mathcal{N}(0,1), \qquad f_\sigma(z) = z, \qquad \varepsilon \sim \mathcal{N}(0, \eta^2),$$

with Z_ρ, Z_σ, and ε mutually independent. Compare the two additive models that share the same nontrivial σ-channel f_σ but differ in the ρ-channel:

$$f_\rho^{(1)}(z) = z, \qquad f_\rho^{(2)}(z) = \frac{z^2 - 1}{\sqrt{2}}.$$

Both satisfy

$$\mathrm{Var}\big(f_\rho^{(1)}(Z_\rho)\big) = \mathrm{Var}\big(f_\rho^{(2)}(Z_\rho)\big) = 1.$$

In both models,

$$R_\rho = X - f_\rho(Z_\rho) = f_\sigma(Z_\sigma) + \varepsilon = Z_\sigma + \varepsilon, \qquad \mathrm{Var}(R_\rho) = 1 + \eta^2,$$

and

$$R_\sigma = X - f_\sigma(Z_\sigma) = f_\rho(Z_\rho) + \varepsilon, \qquad \mathrm{Var}(R_\sigma) = 1 + \eta^2.$$

So the communicated observational residual summaries coincide, but the deterministic shifts are different:

$$m_1(z_\rho, z_\sigma) = z_\rho + z_\sigma, \qquad m_2(z_\rho, z_\sigma) = \frac{z_\rho^2 - 1}{\sqrt{2}} + z_\sigma.$$

Hence, observational residual summaries do not identify m in general. □

C Illustrative Witnesses of Theorem 1

This appendix records two low-dimensional witnesses and one continuous non-Gaussian witness to illustrate the general non-identifiability result in §4. They are concrete illustrations of the same missing interaction direction that the kernel-level proof formalises. The binary case parallels fixed-margin contingency tables (Agresti, 2013) and the broader freedom to vary dependence while preserving marginals (Nelsen, 2006); here the entries are joint-intervention success probabilities, not observational cell probabilities. The Gaussian case illustrates the distinction between zero correlation and independence (Rényi, 1959, Sec. 2): its averaged reports are mixtures and need not be Gaussian. It preserves only their second moments across the family, whereas the binary and Laplace witnesses preserve complete one-cause reports. These connections explain the familiar source of ambiguity while distinguishing the theorem's general kernel-level statement from these illustrations.

C.1 THE BINARY 2×2 WITNESS

Let $X \in \{0,1\}$ and $p_{ij} := Q(\{1\} \mid i,j)$ be its success probability under $\mathrm{do}(Z_\rho = i, Z_\sigma = j)$, for $i,j \in \{0,1\}$. Suppose the other private cause is averaged with weight $\frac{1}{2}$. Write

$$a_i := \tilde{\chi}^\rho(\{1\} \mid i), \qquad b_j := \tilde{\chi}^\sigma(\{1\} \mid j).$$

Then the reported one-cause kernels impose the linear constraints

$$\frac{1}{2}p_{00} + \frac{1}{2}p_{01} = a_0, \qquad \frac{1}{2}p_{10} + \frac{1}{2}p_{11} = a_1,$$
$$\frac{1}{2}p_{00} + \frac{1}{2}p_{10} = b_0, \qquad \frac{1}{2}p_{01} + \frac{1}{2}p_{11} = b_1.$$

This system has rank three, not four, because the sum of the first two equations equals the sum of the second two. Its null space is spanned by

$$(1,-1,-1,1).$$

So if $p_0 = (p_{00}^{(0)}, p_{01}^{(0)}, p_{10}^{(0)}, p_{11}^{(0)})$ is one admissible table with every entry strictly between 0 and 1, then every table of the form

$$p(t) = p_0 + t(1,-1,-1,1)$$

remains a probability table with the same row and column averages for all sufficiently small $|t|$. The unresolved coordinate is exactly an interaction term.

C.2 THE GAUSSIAN COVARIANCE WITNESS

For a moment-based witness, let the shared outcome be vector-valued, $X = (X_1, X_2)$, and let $(Z_\rho, Z_\sigma) \in \{0,1\}^2$. Under each joint intervention cell, define a centred Gaussian response

$$X \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma) \sim \mathcal{N}\big(0, \Sigma_t(z_\rho, z_\sigma)\big),$$

with covariance matrix

$$\Sigma_t(z_\rho, z_\sigma) := \begin{pmatrix} 1 & t(-1)^{z_\rho + z_\sigma} \\ t(-1)^{z_\rho + z_\sigma} & 1 \end{pmatrix}, \qquad |t| < 1.$$

The cell-wise cross-covariance is therefore

$$\sigma_{12}^{(t)}(z_\rho, z_\sigma) := \mathrm{Cov}_t(X_1, X_2 \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)) = t(-1)^{z_\rho + z_\sigma}.$$

If an agent averages over the other private cause with equal weights, then the reported one-cause interventional distribution is a mixture of two Gaussians: $\frac{1}{2}\mathcal{N}(0, \Sigma_t(z_\rho, 0)) + \frac{1}{2}\mathcal{N}(0, \Sigma_t(z_\rho, 1))$, generally non-Gaussian when $t \neq 0$. However, the covariance matrix of this mixture is the equal-weighted average of the cell-wise covariances, so the reported row and column second-moment summaries are

$$\frac{1}{2}\sigma_{12}^{(t)}(z_\rho, 0) + \frac{1}{2}\sigma_{12}^{(t)}(z_\rho, 1) = 0, \qquad \frac{1}{2}\sigma_{12}^{(t)}(0, z_\sigma) + \frac{1}{2}\sigma_{12}^{(t)}(1, z_\sigma) = 0.$$

These summaries vanish because the cross-covariance entries $\sigma_{12}^{(t)}(z_\rho, z_\sigma) = t(-1)^{z_\rho + z_\sigma}$ alternate in sign across the intervention cells: for fixed z_ρ, the values $t(-1)^{z_\rho}$ and $t(-1)^{z_\rho + 1}$ are negatives of each other, so they cancel under equal weighting. Hence, every value of t induces the same *communicated* cross-covariance summaries, even though the joint interventional laws differ cell by cell. In particular,

$$t = 0 \implies X_1 \perp\!\!\!\perp X_2 \qquad \text{and} \qquad t \neq 0 \implies X_1 \not\perp\!\!\!\perp X_2,$$

within each intervention cell. Here, "communicated cross-covariance summaries" refers only to second moments: the equal-weight mixtures coincide across rows and columns for fixed t, but can change as $|t|$ changes, while their covariance matrices remain fixed. Thus this is a moment-level illustration, not the continuum of identical full reports established by the theorem.

C.3 A CONTINUOUS, NON-GAUSSIAN WITNESS

The witnesses in Sections C.1 and C.2 are finite-dimensional. The statement of Theorem 1 applies more broadly, and the construction in Proposition 6 produces an explicit perturbation on any nontrivial measurable outcome space. We illustrate the Theorem's applicability with a continuous outcome and non-Gaussian base kernel.

Let $\mathcal{X} = \mathbb{R}$ with Borel σ-algebra, and let Z_ρ, Z_σ each take values in $[-1, 1]$, with symmetric marginal laws P_{Z_ρ}, P_{Z_σ} admitting densities bounded away from zero. Take the base kernel

$$Q_0(\cdot \mid z_\rho, z_\sigma) = \mathrm{Laplace}\big(\mu(z_\rho, z_\sigma),\, b\big),$$

where $\mu(z_\rho, z_\sigma) = z_\rho + z_\sigma$ is the additive mean and $b > 0$ is fixed. The Laplace base satisfies the local overlap condition in Assumption (A3): on the intervention rectangle $A_\rho \times A_\sigma = [-1, 1]^2$, the mean $\mu(z_\rho, z_\sigma)$ remains in $[-2, 2]$, so the Laplace density is bounded below on any bounded outcome interval. Choose ν as the uniform law on $[-M, M]$ for some $M > 1$ and let $c > 0$ be the corresponding lower bound, so that

$$Q_0(H \mid z_\rho, z_\sigma) \geq c\,\nu(H)$$

for all $(z_\rho, z_\sigma) \in [-1, 1]^2$ and measurable $H \subseteq \mathcal{X}$.

Choose the perturbation seed $s(x) = \mathbf{1}_{[0,1]}(x) - \mathbf{1}_{[-1,0]}(x)$, which is bounded, ν-centred, and not ν-almost-surely zero. The signed measure $S(H) = \int_H s\, d\nu$, in the form of Eq. (10), has $S(\mathcal{X}) = 0$ as in Proposition 6(a) and is supported on $[-1, 1]$. With $u(z_\rho) = z_\rho$ and $v(z_\sigma) = z_\sigma$, the symmetry of the marginal laws verifies the centring part of Assumption (A2). The perturbation family, i.e. Eq. (11) in this witness,

$$Q_t(\cdot \mid z_\rho, z_\sigma) = Q_0(\cdot \mid z_\rho, z_\sigma) + t\, z_\rho\, z_\sigma\, S(\cdot)$$

consists of valid probability kernels for all $|t| < c$. Each Q_t produces the same single-cause marginal kernels as Q_0, but the joint interventional distribution at $(z_\rho^\star, z_\sigma^\star) = (1, 1)$ shifts mass from the negative half-line to the positive half-line in proportion to t. The moments of the single-cause reports coincide across the family, while the joint interventional law differs at any intervention pair with $z_\rho \cdot z_\sigma \neq 0$.

This witness illustrates that the obstruction is not a finite-dimensional or Gaussian artefact: it survives on continuous outcomes, with non-Gaussian base kernels, and extends to heavy-tailed alternatives obtained by replacing the Laplace base with any kernel admitting a local overlap bound on a positive-measure rectangle.

D ADDITIVE SPECIAL CASES UNDER CAUSAL-RESPONSE COMMUNICATION

Recovery from complete single-cause reports. Suppose the additive structural model and the report operators of Section 4 hold, with finite first moments. Put $c_\rho = \int f_\rho\, dP_{Z_\rho}$, $c_\sigma = \int f_\sigma\, dP_{Z_\sigma}$, and $e = \mathbb{E}[\varepsilon]$. The expected outcomes under the two report kernels are

$$a(z_\rho) = f_\rho(z_\rho) + c_\sigma + e, \qquad b(z_\sigma) = c_\rho + f_\sigma(z_\sigma) + e.$$

Their common population average is $\mu = c_\rho + c_\sigma + e$, so

$$a(z_\rho) + b(z_\sigma) - \mu = f_\rho(z_\rho) + f_\sigma(z_\sigma) + e = \mathbb{E}[X \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)].$$

Thus complete causal reports and the specified averaging laws already identify the joint-intervention expectation under separability. Communicating separately identified f_ρ, f_σ is one sufficient way to provide the response information; it is not a necessary additional disclosure when that information is recoverable from the reports. Proposition 4 concerns only residual summaries. The expectation alone does not identify the full kernel. If P_ε is also known, however, subtracting e gives $m(z_\rho, z_\sigma) = a(z_\rho) + b(z_\sigma) - \mu - e$, which determines Q through Eq. (7), as in Remark 5.

For the following two special cases, assume the additive-separable model in Eq. (4). Suppose identified response functions f_ρ, f_σ are available to recipient σ from the received report and its own causal knowledge. The first special case records when identified slopes are enough to identify the deterministic shift m; the second records when the disturbance variance is also recoverable from an observed marginal variance under additional independence assumptions.

Linear additive responses. Suppose further that

$$f_\rho(z_\rho) = \alpha z_\rho, \qquad f_\sigma(z_\sigma) = \beta z_\sigma.$$

Then the identified slopes (α, β) are sufficient to identify

$$m(z_\rho, z_\sigma) = \alpha z_\rho + \beta z_\sigma.$$

Within the linear additive class, they are also necessary because they fully determine the response functions and hence the deterministic shift m. Since $a(z_\rho) - a(0) = \alpha z_\rho$ and $b(z_\sigma) - b(0) = \beta z_\sigma$, the recipient can read these slopes from the received and local kernels as in Example 5.

Residual variance under independence. Suppose in addition that $Z_\rho \perp\!\!\!\perp Z_\sigma$, $\varepsilon \perp\!\!\!\perp (Z_\rho, Z_\sigma)$, and the marginal variance $\mathrm{Var}(X)$ is observed. Then

$$\mathrm{Var}(\varepsilon) = \mathrm{Var}(X) - \mathrm{Var}\big(f_\rho(Z_\rho)\big) - \mathrm{Var}\big(f_\sigma(Z_\sigma)\big).$$

Because $Z_\rho \perp\!\!\!\perp Z_\sigma$ and $\varepsilon \perp\!\!\!\perp (Z_\rho, Z_\sigma)$, the three terms $f_\rho(Z_\rho)$, $f_\sigma(Z_\sigma)$, and ε are pairwise uncorrelated. Therefore

$$\mathrm{Var}(X) = \mathrm{Var}\big(f_\rho(Z_\rho)\big) + \mathrm{Var}\big(f_\sigma(Z_\sigma)\big) + \mathrm{Var}(\varepsilon),$$

identifying the disturbance variance under these additional moment and independence assumptions.

E Further Discussion and Connections

What the local construction establishes. Non-identification of joint effects from separate interventions is established prior work (Saengkyongam & Silva, 2020). Even the existence of a continuum is not, by itself, a stronger identification failure. In Eq. (2), the averaging laws are fixed, so averaging preserves mixtures of kernels. For two compatible kernels Q_1, Q_2 and $0 \le \lambda \le 1$, let $Q_\lambda = (1-\lambda)Q_1 + \lambda Q_2$. This is a Markov kernel and, for $i \in \{\rho, \sigma\}$,

$$\pi^i(Q_\lambda) = (1-\lambda)\pi^i(Q_1) + \lambda \pi^i(Q_2) = \tilde{\chi}^i.$$

Thus two distinct compatible kernels yield a continuum. The operators are linear in the unknown kernel (affine on probability kernels), and the fixed-report equalities define affine constraints. This does not make Q linear in z_ρ, z_σ or assume linear causal mechanisms. The mixture argument concerns the stated kernel class; it need not preserve further restrictions imposed by particular SCMs or abstraction maps. Theorem 1 instead specifies a nonzero direction through *any* baseline satisfying (A1)–(A3), on general measurable spaces, and guarantees admissibility for both signs of a sufficiently small perturbation. More explicitly, the construction has $|S| \le \nu$. With $M = \|u\|_\infty \|v\|_\infty$, it therefore gives

$$\sup_{z_\rho, z_\sigma} \sup_{H} |Q_t(H \mid z_\rho, z_\sigma) - Q_0(H \mid z_\rho, z_\sigma)| \le |t| M.$$

Thus these baselines are not isolated among compatible kernels: arbitrarily small changes can remain invisible to both complete reports. For a witnessing event and intervention pair, write $c = u(z_\rho^\star) v(z_\sigma^\star) S(H^\star) \neq 0$. Its probability changes as $Q_t(H^\star \mid z_\rho^\star, z_\sigma^\star) = Q_0(H^\star \mid z_\rho^\star, z_\sigma^\star) + tc$. This identifies an explicit unresolved direction; it does not give sharp bounds over the entire compatibility set or prove that every policy query changes. No topological or measure-theoretic claim that such baselines are generic is made.

The unknown here is the joint response needed by the shared abstraction. Specifying a network edge does not determine that response, and specifying a compatible response does not by itself construct suitable abstraction maps. Our results isolate the first issue at the kernel level. If reports leave a policy query ambiguous, jointly intervening on the private causes can supply information missing from both reports; observational joint data would require further causal assumptions.

An identified chain contrast. Consider a pure chain $Z_\rho \to Z_\sigma \to X$ with mutually independent exogenous disturbances, so that $Z_\sigma = f_\sigma(Z_\rho, U_\sigma)$ and $X = f_X(Z_\sigma, U_X)$. Suppose the agents share the identified kernels $K_{\rho\sigma}(dz_\sigma \mid z_\rho) = P(dz_\sigma \mid \mathrm{do}(Z_\rho = z_\rho))$ and $K_{\sigma X}(dx \mid z_\sigma) = P(dx \mid \mathrm{do}(Z_\sigma = z_\sigma))$. These must refer to the same population and mediator scale, with the

downstream kernel identified at every mediator value needed for the query. The standard interventional factorisation (Pearl, 2009, Sec. 3.2) then gives

$$P(X \in H \mid \mathrm{do}(Z_\rho = z_\rho)) = \int K_{\sigma X}(H \mid z_\sigma)\, K_{\rho\sigma}(dz_\sigma \mid z_\rho).$$

Indeed, under this intervention the joint law of (Z_σ, X) is $K_{\rho\sigma}(dz_\sigma \mid z_\rho)K_{\sigma X}(dx \mid z_\sigma)$; integrating out Z_σ proves the formula and uniqueness from the two given kernels. If both variables are set, the upstream intervention has no remaining effect on X:

$$Q_{\mathrm{chain}}(H \mid z_\rho, z_\sigma) = P(X \in H \mid \mathrm{do}(Z_\rho = z_\rho, Z_\sigma = z_\sigma)) = K_{\sigma X}(H \mid z_\sigma).$$

Thus the joint-intervention kernel is identified as well. Unlike the common-effect reports, these messages supply the intermediate response and the downstream causal kernel, not two averages of an unknown joint response on X. This is a sufficient identifiable configuration, not a claim that every chain is identified from arbitrary local summaries. In particular, unmeasured confounding between the mediator and outcome can invalidate the first composition formula. The example reinforces the identification requirement: causal structure and adequate reports make it possible to meet it.

Information available in joint-effect identification. Lee et al. (2020) characterise identification from arbitrary surrogate experiments given a common causal graph and the relevant observational or experimental distributions. Our two outcome-only reports omit that richer information. Saengkyongam & Silva (2020) provide the direct negative precedent noted in Section 2; their positive result uses a known acyclic structure and continuous additive-Gaussian-noise SCMs, with pooled observational and single-variable experimental distributions. Their outcome function may contain interactions: additive noise is weaker than the outcome separability used here. Jung et al. (2023) recover joint effects using graphical conditions and conditional responses linked across experiments by measured variables. Those connecting variables have been averaged out of our interface, so their recovery formula cannot in general be evaluated from our reports. These are alternative ways to supply the information or structure missing from separate outcome reports. Work on piecemeal experimentation (Mayo-Wilson, 2019) and on combining experiments for linear cyclic models with latent variables (Eberhardt et al., 2010) instead targets causal structure or mechanisms. Our target is the outcome distribution under each joint intervention; the complete local SCMs can remain private.

Query-specific ambiguity. For a query F, identification means that $\{F(Q) : Q \in \mathcal{Q}_\tau(\tilde{\chi}^\rho, \tilde{\chi}^\sigma)\}$ has one value. Checking this set or computing its range would be a useful diagnostic, but no general procedure is supplied here.

Mutual information between intervention settings and outcomes is a different quantity. Given a chosen kernel Q and a distribution D over which intervention pairs are selected, one can form the joint law $D(dz_\rho, dz_\sigma)Q(dx \mid z_\rho, z_\sigma)$ and compute mutual information when defined. An intervention is a fixed action; D describes the frequencies of actions. If this full joint law were supplied, it would provide information beyond the two reports. Our ambiguity concerns which Q to use in the first place, so neither mutual information nor an information-geometric interpretation automatically resolves it. Query ranges and total-variation diameters are possible measures of disagreement, not new metrics or algorithms established by this paper.

Contingency tables and causal marginals. The binary witness of §C.1 is a causal version of the familiar fact that fixed row and column sums leave one interaction coordinate free in a 2×2 table (Agresti, 2013). More broadly, compatible marginals need not determine a unique joint distribution (Nelsen, 2006). The structural causal marginal problem studies how much of a joint SCM can be recovered from overlapping causal marginals (Gresele et al., 2022). Our theorem studies this ambiguity for the joint-response kernel needed for RCK communication.

Beyond standard observational equivalence. Observational equivalence concerns models that agree on observational information (Peters et al., 2017). Here the models under comparison agree on specified interventional reports as well, yet some joint-intervention outcome laws differ. The compatibility class is defined by those report equalities, rather than by conditional independences or graph orientations. We make no general claim that it is finer or coarser than a Markov equivalence class.

ICA and non-Gaussian identification. Classical ICA shows that a linear mixture of independent non-Gaussian components is identifiable up to scaling and permutation (Comon, 1994). LiNGAM imports that idea into causal discovery: in a linear acyclic SEM with independent non-Gaussian disturbances and no latent confounding, the full causal structure becomes identifiable from observational data because the mixing matrix can be recovered and then aligned with a causal ordering (Shimizu et al., 2006). The broader functional-model literature extends this logic beyond the linear non-Gaussian case to additive-noise and identifiable functional model classes (Hoyer et al., 2008; Peters et al., 2011). Our positive result is different in both target and mechanism. We do not identify a full DAG or mixing matrix from a single joint law, and we do not use non-Gaussianity of disturbances as the identifying lever. Here the target is a joint-intervention kernel. Our sufficient recovery route uses additive separability, identified response functions and a specified disturbance law.

Abstraction, semantic embedding, and SEP. The abstraction literature provides the closest motivational parallel. In work on exact and approximate causal abstraction, the goal is to relate causal models across levels of description (Beckers & Halpern, 2019; Rischel & Weichwald, 2021; Massidda et al., 2023). The semantic embedding principle (SEP) (D'Acunto et al., 2025) is especially relevant as a contrast case: SEP-style abstraction learning asks which abstraction map $\alpha_{\mathcal{X}}$ pushes a low-level model onto a higher-level one and admits a measurable right inverse. In such settings, non-identifiability often appears at the level of the *abstraction map*: multiple maps may be compatible with the same observed low- and high-level distributions.

Our theorem asks a different question. The intervention and outcome meanings are given, and the agents receive two averages of an unknown joint-intervention kernel. Do those reports determine a unique shared response χ^τ? In RCK notation, agreement after restriction requires

$$\alpha_{\mathcal{X}}^{\rho\triangleleft\tau}(\chi^\rho) = \alpha_{\mathcal{X}}^{\sigma\triangleleft\tau}(\chi^\sigma) = \chi^\tau.$$

Several candidate responses χ^τ can reproduce the same reports, so this equality alone does not identify the answer to a joint-intervention query. Given a response and suitable maps, extension gives the recipient's interpretation: $\chi^{\rho,\sigma} = \beta_{\mathcal{X}}^{\sigma\triangleleft\tau}(\chi^\tau)$. Thus abstraction-map learning and response identification concern different unknowns. The common theme is underdetermination by reduced information. In abstraction learning, several maps can relate given models. Here, fixed averaging operators can return the same reports for several kernels. These are different unknowns. In particular, our π^ρ, π^σ act on kernels by integration over intervention inputs; they are not coordinate maps whose pushforwards recover an unspecified joint probability measure. A joint law of actions and outcomes would additionally require an action distribution.

Compositionality. This also connects to work that frames identification compositionally (Friend et al., 2025): valid local causal reports need not determine a unique joint response. Proposition 3 supplies an additive mean restriction, and Remark 5 gives sufficient response information for that structural model. The full kernel additionally follows when its disturbance law is fixed, identified, or communicated. This is one specified way to combine local responses, not a general necessity or a new principle of identification.