Breaking Symmetry: A New Directionality Criterion for Regularity Theories of Causation

Michael Baumgartner and Timm Lampert

Abstract

It is one of the most frequent criticisms of regularity theories of causation that they cannot distinguish between causes and effects based on Boolean determination relations of sufficiency and necessity alone—jeopardizing their goal of reducing causation to Boolean determination. The criticism rests on the fact that determination relations can be symmetric but causation (in acyclic structures) cannot. Regularity theorists have countered by arguing that only simplistic determination structures are symmetric, and by insisting that structures with a minimal complexity of two alternative causes per effect exhibit non-symmetric determination relations. In this paper, we show that this regularity-theoretic response is insufficient to break symmetries of determination and, thus, to identify the direction of causation. To solve this problem, we introduce a new directionality criterion and demonstrate that it suffices to distinguish causes from effects solely on the basis of Boolean determination relations.

1 Introduction

The pioneers of regularity theories of causation, Hume (1748) and Mill (1843), established the direction of causation by definitionally linking it to temporal precedence: causes are defined to precede their effects. This approach, however, fell out of fashion in the mid-20th century, for two reasons. First, it became customary to view the direction of time as being in no less need of clarification than the direction of causation and, moreover, to exploit the latter to identify the former, rather than the other way around (Reichenbach, 1956). Second, cases of simultaneous causation were taken to show that temporal precedence is not necessary for causation (Taylor, 1964); and backward-in-time causation came to be seen as a matter of physical rather than conceptual (im)possibility (Dummett, 1964).

Hence, regularity theorists were left with the challenge of distinguishing causes from effects without relying on temporal precedence. This was one of three critical problems—along with the issues of separating spurious and causal regularities and discerning preempted from genuine causes—that Lewis (1973) considered so severe that he deemed regularity theories beyond repair and abandoned them in favor of his counterfactual theory of causation. The reason, in a nutshell, is that regularity theories aim to analyze (type-level) causation in terms of the Boolean determination relations of sufficiency and necessity between types of events or variables, but while causation is not symmetric unless the structure is cyclic, determination relations can be symmetric even when the underlying causal structure is acyclic. For example, if immersion in water is sufficient and necessary for drowning, then drowning is also sufficient and necessary for immersion in water. But of course, the underlying causal structure is not symmetric: immersion in water causes drowning, not the other way around.

Although Lewis' critique convinced many philosophers of causation, some regularity theorists continued to develop their framework. They met the challenge of distilling the direction of causation from Boolean determination by pointing out that symmetric determination structures tend to be simplistic and by insisting that so-called *collider* structures with a minimal complexity of two alternative causes per effect feature non-symmetric determination relations (Graßhoff and May, 2001; Baumgartner, 2008a). If regularity theorists then assume that real-world causal structures have a complexity that furnishes enough (type-level) colliders to orient determination relations, which seems an innocuous assumption in light of the enormous causal complexity of our world, then they can claim to have solved the directionality problem without temporal precedence. Not only has this proposal convinced many philosophers not directly invested in regularity theories (Psillos, 2010; Andreas and Guenther, 2021), but it is also related to the collider criterion used to distinguish between causes and effects in the framework of causal Bayesian networks (Spirtes et al., 2000, 85) and thus unifies the solutions to the directionality problem across frameworks.

This paper demonstrates that the now standard regularity-theoretic response to the directionality problem is insufficient to consistently break symmetries of determination. We show that there exist acyclic causal structures satisfying the deterministic collider criterion and still featuring symmetric determination relations. Colliders do not reliably identify the direction of causation.

To solve this problem, we introduce a new directionality criterion that consists of two conditions. The first partitions the set of analyzed factors $\mathbf{F}$ into exogenous and endogenous factors. Such a partition is forthcoming if $\mathbf{F}$ contains a unique maximal set of mutually independent factors, a *MaxInd set*. The factors in the MaxInd set are the exogenous ones; the others are endogenous. This condition alone does not solve the directionality problem, as symmetric determination relations can persist among the endogenous factors themselves. Therefore, the second condition requires that the set $\mathbf{F}$ be expanded until every effect Z_i contained in it has a specific determiner Y_{Z_i}, a so-called *specific disjunct*. The intuition behind this condition is that a specific disjunct acts as an independent switch for its effect, analogous to an ideal intervention. This ensures that the causal structure is modular and that the true difference-makers for each effect can be isolated. We end the paper by demonstrating that the combination of a unique MaxInd set and specific disjuncts is sufficient to consistently break all symmetries of determination. It follows that, pace Lewis, it is possible to infer the direction of causation from relations of Boolean determination alone.

Section 2 reviews the state-of-the-art in regularity theories and the deterministic collider criterion. In Section 3, we show that this criterion is not sufficient to identify the direction of causation. Section 4 introduces our new directionality criterion, demonstrates its sufficiency, and embeds it in the regularity-theoretic framework. We conclude in Section 5.

2 Background

2.1 A modern regularity theory of causation

The primary analysandum of modern regularity theories (e.g. Mackie, 1974; Graßhoff and May, 2001; Baumgartner, 2008a; Wright, 2011; Baumgartner and Falk, 2023; Zhang and Zhang, 2023) is the relation of deterministic type-level causation, that is, the relation of causal relevance between types of events, property instantiations, or states of affairs.[1] As

[1]Some regularity-theoretic proposals consider token-level causation to be primary, i.e. causation between token events (e.g. Mackie, 1965), but the criticism raised against these accounts (e.g. Kim, 1971, especially

primary analysans they employ the Boolean determination relations of sufficiency and necessity obtaining in distributions of local matters of fact, which Lewis (1986, xi-x) famously labeled *Humean mosaics*. A Humean mosaic is a matrix of configurations of modally independent properties coincidently instantiated (see e.g. Table 1b).[2] To remain non-committal concerning the ontology of causation, it is common to represent the relata of causal relevance by variables or factors taking values, for instance, A=a_i or B=b_i. In sum, regularity theories analyze "A=a_i is causally relevant to B=b_i" in terms of Boolean determination relations obtaining in Humean mosaics, thereby reducing causal relevance to Humean mosaics.[3]

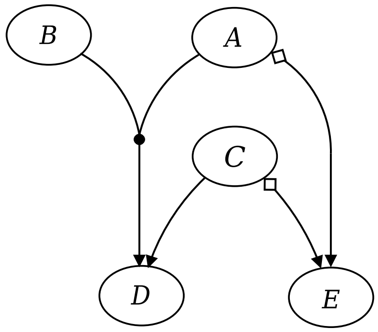
Factors in regularity-theoretic models can represent binary properties or categorical properties with more than two values. For simplicity, we confine ourselves to binary factors in this paper, which lets us abbreviate the explicit "Factor=value" notation. As is customary in Boolean algebra, we write "A" for A=1 and "a" for A=0. Thus, italics are meaningful: "A" designates the factor, while "A" and "a" denote its values, meaning the presence and absence of factor A, respectively. Moreover, we write "$A*B$" for the conjunction "A=1 AND B=1", "$A + B$" for the disjunction "A=1 OR B=1", "$A \rightarrow B$" for the conditional "IF A=1, THEN B=1", and "$A \leftrightarrow B$" for the biconditional "A=1 IFF B=1".

Sufficiency and necessity are given classical (i.e. non-modal) renderings in regularity theories: A is sufficient for B iff $A \rightarrow B$, and A is necessary for B iff $B \rightarrow A$. That a causal structure over a set of factors **F** induces dependencies of sufficiency and necessity is a consequence of two principles assumed by regularity theories (e.g. Graßhoff and May, 2001). The first is *Determinism*. It states that whenever the same complete cause obtains, the same effect occurs; that is, indeterministic associations we commonly observe are due to varying latent factors and not to an indeterministic nature of causation. The second principle is *Causality*, according to which effects do not occur (miraculously) without any of their causes. It follows from these principles that, if the behavior of the factors in **F** is underwritten by a

section IV), in our view, shows that they are beyond repair. Baumgartner (2013) and Andreas and Günther (2024) develop regularity theories of token causation, but they build on primary type-level analyses.

[2]Properties are modally independent if they are logically and conceptually independent and not related in terms of modal dependencies such as supervenience, constitution, grounding, etc.

[3]Like all theories of causation, regularity theories are embedded in a particular metaphysical background that they take for granted. That background is anti-necessitarianism and anti-realism (Hume, 1748, sect. 7), according to which only the causes, effects, and their behavior patterns exist, but not the causal relation itself; there is no causal oomph. Causes do not necessitate or govern their effects, rather, causation and lawhood supervene on Humean mosaics, which are brute facts (see e.g. Beebee, 2000).



(a)

#	A	B	C	D	E
σ_1	1	0	1	1	0
σ_2	1	1	1	1	0
σ_3	0	0	0	0	1
σ_4	1	0	0	0	1
σ_5	0	1	0	0	1
σ_6	1	1	0	1	1
σ_7	0	0	1	1	1
σ_8	0	1	1	1	1

(b)

#	A	C	D	E
σ_1	1	1	1	0
σ_2	1	1	1	0
σ_3	0	0	0	1
σ_4	1	0	0	1
σ_5	0	0	0	1
σ_6	1	0	1	1
σ_7	0	1	1	1
σ_8	0	1	1	1

(c)

Figure/Table 1: An exemplary causal structure (a) where arrows represent causal relevance, "•" symbolizes conjunction, and "⋄" negation, with a corresponding complete Humean mosaic (b) and an incomplete one (c).

causal structure, complete causes are individually sufficient and their disjunction necessary for their effects in **F**. In other words, it is *necessary* for the truth of "A is causally relevant to B" that A is part of some (complex) sufficient and necessary condition of B.

But of course, that A is part of a sufficient and necessary condition of B is *not sufficient* for the truth of "A is causally relevant to B". Most Boolean dependencies correspond to mere patterns of co-occurrence. To show this, suppose that the true causal structure over the binary factors in $\mathbf{F}_1 = \{\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}, \mathrm{E}\}$ is the one in Figure 1a. This causal hypergraph has two non-standard elements: arrows merged by "•" symbolize conjunctive relevance, and "⋄" expresses that the negation of the factor at the tail of the arrow is relevant. That is, in Figure 1a, $A{*}B$ and C are two alternative causes of D and a and c are two alternative causes of E.[4]

If we assume that there are no (latent) causal paths to D and E other than those through A, B, and C, it follows that the factors in $\mathbf{F}_1$ can be combined in exactly the eight types of configurations σ_1 to σ_8 in Table 1b. There are three *exogenous* factors whose values do not have causes in the structure of Figure 1a: A, B, and C. They can be combined in all ($2^3 = 8$) logically possible configurations, each of which is contained in Table 1b. Factors D and E are *endogenous*, meaning their values are caused by values of other factors in accordance

[4]To help with intuitions, the structure can be interpreted to represent the power supply in a city with two power stations: a wind farm and a nuclear plant. A expresses that the wind farm is operational and C that the nuclear plant is operational, and let operationality be sufficient for a nuclear plant to produce electricity, while a wind farm produces electricity provided it is operational and there is wind (B). Hence, the wind farm being operational while it is windy or the nuclear plant being operational ($A{*}B + C$) are two alternative causes of the city being power supplied (D). Whereas the wind farm or the nuclear plant not being operational ($a + c$) are two alternative causes of an alarm being triggered (E).

with the structure in Figure 1a. Table 1b contains a complete distribution of matters of fact; it is the *complete Humean mosaic* for that causal structure.

Now, let us examine some of the Boolean dependencies in that mosaic. For example, $\neg(a+B+c+E)$ is sufficient for D because the mosaic does not feature a row where $\neg(a+B+c+E)$ is true and D false. But as negated disjunctions are not directly causally interpretable, it has become customary to constrain the expressions considered by regularity theories to disjunctive normal forms (DNFs) (Mackie, 1974). If we apply De Morgan's laws to $\neg(a+B+c+E)$, we get $A{*}b{*}C{*}e$, which is in DNF and also sufficient for D. Nowhere in Table 1b is $A{*}b{*}C{*}e$ combined with d. The same holds for $A{*}B{*}C{*}e$, $A{*}B{*}c{*}E$, etc. Moreover, the disjunction of all these sufficient conditions is necessary for D; that is, the following relations of sufficiency and necessity (in DNF) obtain among D and the other factors in $\mathbf{F}_1$:

$$A{*}b{*}C{*}e \ + \ A{*}B{*}C{*}e \ + \ A{*}B{*}c{*}E \ + \ a{*}b{*}C{*}E \ + \ a{*}B{*}C{*}E \ \leftrightarrow \ D \tag{1}$$

As (1) is true for all the configurations in Table 1b, (1) is *entailed* by that mosaic. Moreover, by expressing D as a function of the other factors in $\mathbf{F}_1$, (1) perfectly accounts for D. But (1) does not track causation. E, for example, appears in various sufficient conditions of D, which in turn are part of a necessary condition of D, yet E is not causally relevant for D. Still, some relations of sufficiency and necessity reflect causation; in Table 1b, $A{*}B$ and C are individually sufficient and their disjunction is necessary for D:

$$A{*}B \ + \ C \ \leftrightarrow \ D \tag{2}$$

(2) is also entailed by Table 1b, it also perfectly accounts for D, and, unlike (1), it is directly causally interpretable from left to right: $A{*}B$ and C are the two alternative causes of D. Accordingly, the crucial problem to be solved by a regularity theory is to filter out those Boolean dependencies obtaining in Humean mosaics that track causation.

The principle exploited for that purpose is *redundancy-freeness*. If a Boolean function, such as (1), that is entailed by a mosaic $\mathcal{M}$ and accounts for the behavior of (some of) the factors in $\mathcal{M}$ contains elements that can be removed from it such that the reduced function, for example (2), is also entailed by $\mathcal{M}$ and equally accounts for the (same) factors, the removed elements make no difference. What is not a difference-maker is not part of the causal structure. Therefore, filtering out the causally interpretable sufficiency and necessity relations requires eliminating redundant elements.

Boolean functions may contain redundancies along four dimensions, and respective redundancy-freeness constraints have been gradually added to regularity theories over time. Mackie (1974) required that sufficient conditions must be minimal, that is, they must not contain conjuncts that can be removed without the remaining condition losing sufficiency. A condition such as $A{*}B$ that is sufficient for D is *minimally sufficient* iff none of its proper parts (A and B) are sufficient for D. In the mosaic of Table 1b, that holds for $A{*}B$ but not, say, for $A{*}b{*}C{*}e$ because its proper part C is itself sufficient for D. Hence, Mackie (1974) stipulated that only necessary disjunctions of minimally sufficient conditions (in DNF) track causation.

But Mackie's theory could not distinguish between spurious and causal regularities. Therefore, Graßhoff and May (2001) introduced the constraint that necessity relations must be minimized as well. A necessary disjunction of minimally sufficient conditions, such as $A{*}B + C$ in (2), is *minimally necessary* iff no disjuncts can be removed from it without the remainder losing necessity. That holds for the left-hand side of (2) because neither $A{*}B$ nor C is necessary for D. Only minimally necessary disjunctions of minimally sufficient conditions (in DNF) track causation. Biconditionals such as (2) with redundancy-free DNFs on the left-hand side are called *RDN-biconditionals* (for 'Redundancy-free Disjunctive-Normal-form biconditionals'):

RDN-biconditional: A biconditional $\Phi \leftrightarrow Z$ entailed by a mosaic $\mathcal{M}$ is an RDN-biconditional if, and only if, according to $\mathcal{M}$, Φ is a minimally necessary disjunction, with at least two disjuncts,[5] of minimally sufficient conjunctions in DNF and Z is a single factor value.

Combining all RDN-biconditionals entailed by a Humean mosaic to a complex Boolean function may give rise to a third dimension of redundancies. For example, Table 1b entails three RDN-biconditionals. The following is the resulting complex function:

$$(A{*}B + C \leftrightarrow D){*}(a + c \leftrightarrow E){*}(a{*}D + e \leftrightarrow C) \tag{3}$$

[5] This is the collider criterion that will be explained in section 2.2 and dropped in section 4.

Even though all sufficiency and necessity relations in (3) are minimized, (3) does not reflect causation because $a{*}D$ and e are not two alternative causes of C in Figure 1a. It turns out, however, that (3) is logically equivalent to (4):

$$(A{*}B + C \leftrightarrow D){*}(a + c \leftrightarrow E) \tag{4}$$

That is, $(a{*}D + e \leftrightarrow C)$ can be removed from (3) such that the remainder equally accounts for the behavior of the factors in Table 1b. $(a{*}D + e \leftrightarrow C)$ is structurally redundant. A conjunction of RDN-biconditionals entailed by a mosaic $\mathcal{M}$ is causally interpretable only if that conjunction is *structurally minimal*, meaning that no RDN-biconditional can be removed from it while logical equivalence is preserved (Baumgartner and Falk, 2023).

Finally, Zhang and Zhang (2023) identify a fourth dimension of redundancies. It is possible that a mosaic $\mathcal{M}$ entails multiple RDN-biconditionals with the same factor value on the right-hand side. Let $\Phi_1 \leftrightarrow Z$ and $\Phi_2 \leftrightarrow Z$ be two such RDN-biconditionals, and let $\mathbf{F}_{\Phi_1}$ and $\mathbf{F}_{\Phi_2}$ be the sets of factors appearing in Φ_1 and Φ_2, respectively. If one of $\mathbf{F}_{\Phi_1}$ or $\mathbf{F}_{\Phi_2}$ is a proper subset of the other, one of $\Phi_1 \leftrightarrow Z$ or $\Phi_2 \leftrightarrow Z$ accounts for Z with at least one fewer factor—let us call it R—than the other. Factor R then is redundant because Z can be expressed as a Boolean function without R. Hence, Zhang and Zhang (2023) require that a causally interpretable RDN-biconditional $\Phi \leftrightarrow Z$ must be *functionally minimal*, meaning that $\mathbf{F}_\Phi$ does not contain a proper subset $\mathbf{F}'_\Phi$ such that Z can be expressed as a Boolean function of $\mathbf{F}'_\Phi$.[6] Zhang and Zhang (2023, 13) also show that there is a tight connection between functional minimality and difference-making: a mosaic $\mathcal{M}$ that entails a functionally minimal RDN-biconditional $\Phi \leftrightarrow Z$ is guaranteed to contain, for every factor in $\mathbf{F}_\Phi$, two rows in which that factor and Z vary while all other factors in $\mathbf{F}_\Phi$ remain constant. That is, every factor in Φ is indispensable to account for the variation of Z in at least two rows of $\mathcal{M}$.[7]

[6] Functional minimality differs from Graßhoff and May's (2001) minimal necessity requirement: the latter eliminates redundant disjuncts from a necessary condition, whereas the former eliminates redundant factors from the factor set on the basis of which the outcome is modeled (see Zhang and Zhang, 2023, for a detailed discussion of this difference).

[7] Despite this connection between functional minimality and difference-making, one may argue that functional minimality excludes certain well-behaved causal structures and, thus, should not be imposed as a necessary constraint on causal relevance. For the sake of our argument, however, we presume functional minimality in the following because we intend to show that a collider criterion does not suffice to identify the direction of causation even on the basis of the strictest criteria of redundancy-freeness. De Souter (2025) recently defended

A structure of Boolean determination that is redundancy-free along all four of the above dimensions is called a *minimal theory* (Baumgartner and Falk, 2023).

Minimal theory: A minimal theory for a Humean mosaic $\mathcal{M}$ is a structurally minimal conjunction (with one or more conjuncts) of functionally minimal RDN-biconditionals entailed by $\mathcal{M}$.

One last issue requires attention before we can define causal relevance. Complete redundancy elimination along the four redundancy dimensions is relative to the analyzed factor set. That is, factor values contained in a minimal theory relative to a set $\mathbf{F}_i$ may not be contained in a minimal theory relative to a superset $\mathbf{F}_j \supset \mathbf{F}_i$. To demonstrate this, we re-analyze the causal structure in Figure 1a based on $\mathbf{F}_2 = \{\mathrm{A, C, D, E}\}$, which is equal to $\mathbf{F}_1 \setminus \{\mathrm{B}\}$. Relative to $\mathbf{F}_2$, that structure yields the mosaic in Table 1c. This is an *incomplete* Humean mosaic that does not represent one causal path to D: the one through B. As a result, D cannot be expressed as a Boolean function of $\mathbf{F}_2 \setminus \{\mathrm{D}\}$, for in rows σ_4 and σ_6 all factors in $\mathbf{F}_2 \setminus \{\mathrm{D}\}$ are constant while D changes. The following is the minimal theory entailed by Table 1c:

$$(a + c \leftrightarrow E) * (a * D + e \leftrightarrow C) \tag{5}$$

However, C is not an effect in Figure 1a. Thus, (5) does not track causation. The reason why $(a * D + e \leftrightarrow C)$ is not identifiable as spurious is that there exists a latent causal path inducing a variation in the factors of $\mathbf{F}_2$ that cannot otherwise be accounted for. But when $\mathbf{F}_2$ is expanded to $\mathbf{F}_1$, no varying latent paths remain, and $(a * D + e \leftrightarrow C)$ is no longer needed to account for the behavior of the factors in $\mathbf{F}_1$; and the resulting minimal theory is (4), which tracks causation. More generally, Boolean determination relations that appear to be of a difference-making type relative to a set $\mathbf{F}$, but actually are spurious, are identified as such in the course of gradually expanding $\mathbf{F}$. Accordingly, modern regularity theories complement conditions securing redundancy-freeness with a permanence constraint. A minimal theory Ψ for a mosaic $\mathcal{M}$ relative to a set $\mathbf{F}$ is causally interpretable only if it is *permanently redundancy-free*, meaning there is no expansion $\mathbf{F}^*$ of $\mathbf{F}$, $\mathbf{F}^* \supset \mathbf{F}$, with corresponding mosaic $\mathcal{M}^*$ such that the minimal theory Ψ^* entailed by $\mathcal{M}^*$ reveals redundancies in Ψ, because not

the incorporation of functional minimality into a regularity-theoretic definition of causation.

all components of Ψ are also components of Ψ^*.

Now we can define causal relevance (type-level causation) in regularity-theoretic terms.[8]

Causal Relevance (CR): A is causally relevant for B if, and only if, the factors A and B are part of a Humean mosaic $\mathcal{M}$ that entails a permanently redundancy-free minimal theory Ψ comprising an RDN-biconditional $\Phi \leftrightarrow B$, such that A is a part of Φ.

It is important to regularity theorists that their definition of causal relevance be directly applicable to data in causal modeling contexts. Data can be treated as Humean mosaics, minimal theories can be inferred from them, and the permanence constraint can be checked across a certain number of expansions until the evidence is considered sufficient for an inductive leap to accepting permanent redundancy-freeness. The present paper, however, is concerned only with the conceptual question whether a regularity theory can, in principle, distinguish between causes and effects, that is, whether it can do so when Humean mosaics are ideal. More specifically, the question we pursue in the following is whether a regularity theory can orient causal relations on the basis of complete Humean mosaics, in which the causally relevant regularity structure is fully represented, no causal paths remain latent relative to the analyzed factor set, and no possible configurations of the factors are missing.

2.2 Deterministic colliders

This section introduces the collider criterion currently used to distinguish between causes and effects in regularity theories in the vein of CR. We begin by narrowing our discussion to *acyclic* causal structures. The reason is that, in causal cycles, some symmetric determination relations are to be expected, and a comprehensive distinction between (type-level) causes and effects is not forthcoming. That is not to say that questions of directionality cannot also arise for cyclic structures, but in this paper, we restrict attention to the problem of orienting causal relevance where the underlying structure is itself non-symmetric. Hence, in what follows, we shall discuss acyclic structures exclusively, without repeatedly qualifying them as such.

The previous section showed that CR has no difficulty distinguishing causes from effects based on the determination relations entailed by the mosaic in Table 1b. Standard wisdom

[8] This definition is inspired by Baumgartner and Falk (2023). But their definition is more general; it defines causal relevance not just for single factors values as A but for Boolean expressions in DNF.

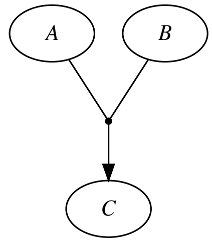
holds that this generalizes beyond this example. The idea is that CR succeeds in identifying the direction of causation without reliance on time because RDN-biconditionals in minimal theories are defined to have at least two disjuncts (see, e.g., Graßhoff and May, 2001; Baumgartner, 2008a; Psillos, 2010; Andreas and Guenther, 2021). To understand the rationale behind this, consider the RDN-biconditional $a + c \leftrightarrow E$ entailed by Table 1b. Each of a and c determines E, but E determines neither a nor c. Instead, E only determines the disjunction $a + c$, which cannot be causally interpreted in line with the principle of Determinism. Determinism requires that whenever the same cause is present, the same effect occurs. Yet in Table 1b, when E is given, sometimes a without c obtains, sometimes c without a, and sometimes both a and c obtain.

It might be argued that, although E alone neither determines a nor c, E is still part of a minimally sufficient condition for each of a and c. Whenever E is present and c is not—meaning that C obtains—it follows that a is present (because if E is not caused by c, it must be caused by a according to Figure 1a). In other words, $E{*}C$ is minimally sufficient for a and thus determines a. Analogously, $E{*}A$ is minimally sufficient for c. However, neither $E{*}C \rightarrow a$ nor $E{*}A \rightarrow c$ can be causally interpreted in line with the principle of Causality, as Table 1b contains two configurations, σ_3 and σ_5, featuring a and c without $E{*}C$ and $E{*}A$. That is, $E{*}C$ and $E{*}A$ are not necessary for a and c, respectively; and they cannot be disjunctively complemented to necessary conditions of a and c by additional minimally sufficient conditions either.[9] By contrast, a and c not only individually determine E, but their disjunction is also necessary for E. That is, $a + c \leftrightarrow E$ can be causally interpreted from left to right in accordance with both Determinism and Causality, whereas E is not part of an analogous dependency structure accounting for a or c. The regularity patterns among a, c, and E are not symmetric; rather, they entail that a and c are the causes and E is the effect.

E is a deterministic collider:

Deterministic Collider: A factor value Z is a deterministic collider in a Humean mosaic $\mathcal{M}$ if, and only if, $\mathcal{M}$ entails at least two minimally sufficient conditions for Z whose disjunction is minimally necessary for Z in $\mathcal{M}$.

[9] To illustrate for a. The set of all minimally sufficient conditions for a in Table 1b is $\{E{*}C, B{*}d, b{*}D{*}E\}$. However, $E{*}C + B{*}d + b{*}D{*}E$ is not necessary for a because in σ_3 a is given without any of these conditions. For analogous reasons, the disjunction of all minimally sufficient conditions of c is not necessary for c.



(a)

#	A	B	C
σ_1	1	1	1
σ_2	0	1	0
σ_3	1	0	0
σ_4	0	0	0

(b)

Figure/Table 2: A simple conjunctive cause (a) with its corresponding Humean mosaic (b).

Deterministic colliders are widely considered sufficient for breaking symmetries of determination because, under Determinism and Causality, they can only be causally interpreted in one direction. This approach to symmetry breaking, if valid, can be readily generalized to all minimal theories, as all effects in minimal theories are deterministic colliders. The reason is that minimal theories are conjunctions of RDN-biconditionals, which are defined to have at least two disjuncts. If causal relevance is then defined as in CR, it follows that all effects can be distinguished from their causes without relying on temporal precedence.

It is important to note that, while commonly assumed to be sufficient for the identifiability of the direction of causation, deterministic colliders are not necessary for that purpose. To see this, consider the very simple mosaic in Table 2b. Factors A and B vary independently of each other and C takes the value 1 if, and only if, both A and B are 1:

$$A{*}B \ \leftrightarrow C \tag{6}$$

$A{*}B$ is minimally sufficient and minimally necessary for C in Table 2b, but although C is likewise minimally sufficient for A and for B, Table 2b does not feature a necessary condition for either A or B. In configurations σ_3 and σ_4, factor A varies while B and C remain constant, and in σ_2 and σ_4 B varies without variation in the other factors. This shows that A and B cannot be effects in a deterministic structure that would satisfy the principle of Causality. The only factor whose values can be modeled as effects in accordance with both Determinism and Causality is C. Hence, even though C is no deterministic collider, the relations of Boolean determination encoded in (6) are not symmetric.

The task of formulating a criterion that is necessary for identifying the direction of causation has not been tackled by regularity theorists to date. To meet the Lewisian challenge of distinguishing between causes and effects without time, it is enough to have a criterion that is sufficient for breaking determination symmetries. Current wisdom holds that the following meets this requirement: if Z is a deterministic collider in a minimal theory Ψ, then Z is an effect and its two or more minimally sufficient conditions in Ψ are causes. Although deterministic colliders are not necessary for symmetry breaking as such, a regularity theory that relies on the collider criterion as its general route to directionality must assume that all type-level effects fall within the criterion's scope, meaning that they have at least two alternative causes.[10] This assumption is not too costly, for regularity theorists do not analyze causation for all possible (toy) worlds but for our actual, causally immensely complex world, which is very likely to furnish two alternative causes for every type of effect. They therefore accept the corollary that only worlds of a certain complexity feature causation.

Before the next section critically discusses this directionality criterion, let us give it a more precise formulation. In what follows, we use superscripts in square brackets, as in $\Phi^{[Z_i]} \leftrightarrow Z_k$, to indicate that Z_i appears on the left-hand side of the relevant RDN-biconditional.

Unambiguous Orientability: A Humean mosaic $\mathcal{M}$ is unambiguously orientable if, and only if, $\mathcal{M}$ does not entail two minimal theories Ψ_1 and Ψ_2 such that Ψ_1 comprises an RDN-biconditional $\Phi_i^{[Z_1]} \leftrightarrow Z_2$ and Ψ_2 comprises an RDN-biconditional $\Phi_k^{[Z_2]} \leftrightarrow Z_1$.

If $\mathcal{M}$ does entail two such minimal theories, there are factor values Z_1 and Z_2 for which it remains unclear whether Z_1 is causally relevant for Z_2 or vice versa. If it does not, all causal relations entailed by $\mathcal{M}$ can only be oriented in one direction.[11]

Here, then, is a more precise formulation of the collider criterion:

[10]Borrowing terminology from Weinberger et al. (2024), it may be said that this minimal complexity assumption is part of CR's *worldly infrastructure of causation*. Similarly, an interventionist theory assumes that, outside of a modeled system, there exists a possible intervention on every variable within that system (Woodward, 2003).

[11]Two remarks may be helpful here. First, the problem of unambiguous orientability differs from what Baumgartner (2008b) calls the *causal chain problem*. The chain problem concerns the empirical indistinguishability of two minimal theories, one representing a causal chain and the other a common-cause structure. These theories do not, however, imply opposed causal orientations. By contrast, the problem of unambiguous orientability arises when one and the same mosaic entails two minimal theories that do imply opposed orientations. Second, we will treat the trivial case of a mosaic entailing no minimal theories at all as unambiguously orientable. That is, the empty orientation is unambiguous.

Collider Directionality (CD): If all effects contained in a Humean mosaic $\mathcal{M}$ are deterministic colliders and $\mathcal{M}$ entails a minimal theory Ψ_1 comprising an RDN-biconditional $\Phi_i^{[Z_1]} \leftrightarrow Z_2$, then $\mathcal{M}$ does not entail another minimal theory Ψ_2 comprising an RDN-biconditional $\Phi_k^{[Z_2]} \leftrightarrow Z_1$.

3 Symmetries in deterministic colliders

This section shows that all effects being deterministic colliders does not ensure that the direction of causation is identifiable from a Humean mosaic. We demonstrate this by example. Consider the mosaic in Table 3. It entails three minimal theories, (7) to (9), for the presence of factors A, B, and C, and three minimal theories, (10) to (12), for their absence. For brevity, we call the former the *positive* and the latter the *negative* minimal theories. All of them are mutually logically equivalent; they are true for exactly the configurations in Table 3; and the positive and negative theories are transferrable into one another using De Morgan's laws. Listing all of them is therefore unnecessary. We do so only to exhibit the complete symmetry of the dependence relations. Henceforth, we focus on the positive minimal theories only.

The symmetries in our example demonstrate that every factor can be expressed as a Boolean function of the other two factors. For every pair of factors $\langle \mathrm{Z}_i, \mathrm{Z}_j \rangle$ in Table 3 it holds that Z_i and Z_j are configured in all logically possible combinations, allowing them to be modeled as exogenous and leaving the third factor Z_k as endogenous. But despite the direction of causation being unclear, all candidate effects Z_k are deterministic colliders. Hence, there exist mosaics meeting the requirement that all effects are deterministic colliders and yet failing unambiguous orientability: they entail two minimal theories such that one

#	A	B	C
σ_1	1	1	1
σ_2	0	0	1
σ_3	0	1	0
σ_4	1	0	0

$$B{*}C + b{*}c \leftrightarrow A \tag{7}$$
$$A{*}C + a{*}c \leftrightarrow B \tag{8}$$
$$A{*}B + a{*}b \leftrightarrow C \tag{9}$$
$$b{*}C + B{*}c \leftrightarrow a \tag{10}$$
$$a{*}C + A{*}c \leftrightarrow b \tag{11}$$
$$a{*}B + A{*}b \leftrightarrow c \tag{12}$$

Table 3: A Humean mosaic with all six minimal theories it implies.

comprises an RDN-biconditional $\Phi_i^{[Z_1]} \leftrightarrow Z_2$ and the other an RDN-biconditional $\Phi_k^{[Z_2]} \leftrightarrow Z_1$. Collider Directionality (CD) is false.

It should be emphasized that this is not an artificially constructed example; rather, commonplace processes can generate mosaics as in Table 3. Consider a staircase light, modeled by factor C with values "on" and "off", controlled by two switches (one at the top and one at the bottom of the staircase) represented by factors A and B with values "up" and "down". The light is on (C) iff both switches are either in the up ($A{*}B$) or in the down position ($a{*}b$). Alternatively, factor A can be interpreted to stand for "high" and "low" skills that people may have for a given activity, factor B for "high" and "low" challenges posed by that activity, and C for the "presence" and "absence" of the autotelic experience of complete involvement with the activity called *flow*. According to Csikszentmihalyi (1975, ch. 4), flow occurs, C, iff skills and challenges are either both high ($A{*}B$) or both low ($a{*}b$). If Table 3 stems from one of these structures, the minimal theory that correctly captures the direction of causation is (9) (and (12), respectively). However, nothing in the dependencies encoded in that mosaic identifies C as the effect.

Table 3 provides the simplest type of counterexample to CD. If (7) to (9) are embedded in more complex minimal theories, their symmetries carry over to the resulting complex structures. To illustrate, the eight minimal theories in Table 4 are all logically equivalent, meaning they are entailed by (and entail) the exact same mosaic. In all of them, exactly two factors are endogenous, but it is a different pair of factors in each—despite all candidate effects being deterministic colliders. Thus, CD fails for a wide range of complex structures.

The symmetries in Tables 3 and 4 are so comprehensive that distinguishing between exogenous and endogenous factors is impossible. But even in cases where such a distinction can be made, orienting all causal relations may remain unfeasible because among the en-

$(b{*}c + B{*}C \leftrightarrow A){*}(d{*}E + D{*}e \leftrightarrow C)$	$(a{*}c + A{*}C \leftrightarrow B){*}(d{*}E + D{*}e \leftrightarrow C)$
$(b{*}c + B{*}C \leftrightarrow A){*}(c{*}E + C{*}e \leftrightarrow D)$	$(a{*}c + A{*}C \leftrightarrow B){*}(c{*}E + C{*}e \leftrightarrow D)$
$(b{*}c + B{*}C \leftrightarrow A){*}(c{*}D + C{*}d \leftrightarrow E)$	$(a{*}c + A{*}C \leftrightarrow B){*}(c{*}D + C{*}d \leftrightarrow E)$
$(A{*}B + a{*}b \leftrightarrow C){*}(c{*}E + C{*}e \leftrightarrow D)$	$(A{*}B + a{*}b \leftrightarrow C){*}(c{*}D + C{*}d \leftrightarrow E)$

Table 4: Eight logically equivalent minimal theories entailing perfectly symmetric determination relations.

#	A	B	C	D	E
σ_1	1	0	1	1	1
σ_2	1	1	1	0	1
σ_3	0	1	1	0	1
σ_4	1	1	0	0	1
σ_5	0	1	0	0	1
σ_6	1	0	0	1	0
σ_7	0	0	0	1	0
σ_8	0	0	1	0	0

$$(A{*}b \; + \; b{*}c \;\; \leftrightarrow \;\; D){*}(B \; + \; C{*}D \;\; \leftrightarrow \;\; E) \qquad (13)$$
$$(E{*}b \; + \; b{*}c \;\; \leftrightarrow \;\; D){*}(B \; + \; A{*}C \;\; \leftrightarrow \;\; E) \qquad (14)$$
$$(E{*}b \; + \; c{*}e \;\; \leftrightarrow \;\; D){*}(B \; + \; A{*}C \;\; \leftrightarrow \;\; E) \qquad (15)$$

Table 5: A Humean mosaic with three minimal theories it implies; red is used for highlighting only.

dogenous factors there are two factors Z_1 and Z_2 for which it is unclear whether the values of Z_1 cause those of Z_2, or vice versa. The mosaic in Table 5 provides an illustration. It entails the three minimal theories (13) to (15), all of which agree that A, B, and C are exogenous, while D and E are endogenous. However, according to (13), D is a cause of E, whereas in (14) and (15), E is a cause of D. That is, despite a clear distinction between exogenous and endogenous factors, the direction of causation between D and E remains indeterminate. Yet again, all effects in (13) to (15) are deterministic colliders.

It is evident that something is amiss with the standard approach to identifying the direction of causation within the regularity-theoretic framework. The next section will first diagnose the core of the problem and then propose a solution.

4 A new directionality criterion

Readers familiar with the literature on causal Bayesian networks, where colliders are prominently used as empirical markers of the direction of causation, might suspect that the reason deterministic colliders, as defined in Section 2.2, are insufficient to identify the direction of causation is that the minimally sufficient conditions constituting deterministic colliders are not required to be *independent* of each other. In all minimal theories entailed by the mosaic in Table 3, the disjuncts are mutually exclusive (i.e., the presence of one disjunct entails the absence of the other), which is a strong form of dependence. In Bayesian networks, however, only so-called *unshielded* colliders have a determinate orientation. An unshielded collider is a triple of variables $\langle Z_1, Z_2, Z_3 \rangle$ such that Z_1 and Z_3 can be rendered independent by conditioning on a set of variables not containing Z_2, both are correlated with Z_2, and they

#	A	B	D	C
σ_1	0	1	1	1
σ_2	1	0	1	1
σ_3	0	0	1	1
σ_4	1	0	0	1
σ_5	1	1	1	0
σ_6	1	1	0	0
σ_7	0	1	0	0
σ_8	0	0	0	0

$$A{*}b \;+\; a{*}D \;\leftrightarrow\; C \tag{16}$$

Table 6: A Humean mosaic that allows for identifying the direction of causation, and its corresponding minimal theory.

become dependent when conditioned on Z_2. This dependency pattern has only one viable causal orientation: $Z_1 \longrightarrow Z_2 \longleftarrow Z_3$ (Spirtes et al., 2000, 85).

The lack of independent variability of the minimally sufficient conditions in the example of Table 3 is part of the reason for the failure of CD, but it is not the heart of it. In fact, deterministic colliders with mutually dependent minimally sufficient conditions can often be causally oriented. A case in point is the mosaic in Table 6 with (16) being the only minimal theory it entails. Here, C is a deterministic collider whose two minimally sufficient conditions are mutually exclusive: one condition features A, the other a. Yet, despite the dependence between these conditions, the determination relations in Table 6 admit only the one causal orientation given by (16). The key difference between the mosaics in Tables 3 and 6 is that, in the former, *every* change of any factor is accompanied by a change in at least one other factor, whereas three factors—A, B, and D—have at least two rows in the latter mosaic in which they are the only varying factors.[12]

The following is a sufficient condition for exogeneity: a factor Z is exogenous in a Humean mosaic $\mathcal{M}$ if $\mathcal{M}$ contains two rows in which Z changes its value while all other factors remain constant. Such a Z cannot be expressed as a Boolean function of the other factors in $\mathcal{M}$. Conversely, if Z never changes in isolation, it does not follow that Z is endogenous (i.e. not exogenous). Table 3 is a case in point: no factor satisfies the exogeneity condition, yet it does not follow that all factors are endogenous; it only follows that all factors

[12]In other words, the Hamming distance between any two rows of Table 3 is 2, while for every row in Table 6 there is at least one other row at a Hamming distance of 1 (for more on Hamming distance see Mano and Ciletti, 2012).

could be modeled as endogenous, leaving it ambiguous which ones actually are endogenous. Failure of the exogeneity condition is merely a necessary condition for endogeneity: a factor Z is endogenous in $\mathcal{M}$ only if $\mathcal{M}$ has no two rows in which Z changes while all other factors remain constant. This allows Z's behavior to be expressed as a Boolean function of the other factors. Jointly, the exogeneity and endogeneity conditions leave substantial room for ambiguity, especially in mosaics like Table 3, where too many factors satisfy the latter condition and too few the former.

To partition the set of analyzed factors $\mathbf{F}$ into exogenous and endogenous ones, there must be exactly one split whose first component meets the exogeneity condition and whose complement meets the endogeneity condition. We formalize this requirement by introducing a set-level counterpart of the above exogeneity condition, that is, a condition that tests the joint independence of a candidate set of factors. In a complete Humean mosaic, the exogenous factors jointly realize all logically possible configurations; they form a maximal set of mutually independent factors, or a *MaxInd set*.

MaxInd set: A non-empty set of factors $\mathbf{F}_{\text{sub}} \subseteq \mathbf{F}$ is a *MaxInd set* for the Humean mosaic $\mathcal{M}$ if, and only if, it meets both of the following:

(1) $\mathcal{M}$ contains all 2^n logically possible configurations of the n factors in $\mathbf{F}_{\text{sub}}$.

(2) $\mathbf{F}_{\text{sub}}$ is maximal, meaning that no factor $Z \in \mathbf{F} \setminus \mathbf{F}_{\text{sub}}$ can be added to $\mathbf{F}_{\text{sub}}$ without violating condition (1).

This notion provides a regularity-theoretic rendering of the intuition that the exogenous factors are the mutually independent ones without causes in the modeled system. For the partition of $\mathbf{F}$ to be unambiguous, the complement of the MaxInd set must satisfy the endogeneity condition, meaning that none of its members changes its value in isolation, and there must be only one such split. We express these requirements in the following condition:

Unique MaxInd Set Condition (UMS): A Humean mosaic $\mathcal{M}$ satisfies UMS if, and only if, $\mathcal{M}$ has a unique MaxInd set $\mathbf{F}_{\text{exo}} \subset \mathbf{F}$ such that all remaining factors in $\mathbf{F}_{\text{end}} = \mathbf{F} \setminus \mathbf{F}_{\text{exo}}$ do not change their values in isolation in $\mathcal{M}$.

Two clarifications are important. First, satisfying UMS does not entail that some values of the exogenous and endogenous factors are actually causally related. It only fixes which factors count as exogenous and which count as endogenous in the mosaic. Second, satisfying UMS is not sufficient to break all possible symmetries among endogenous factors. A mosaic satisfying UMS may contain multiple endogenous factors whose causal relations are not unambiguously orientable. The mosaic in Table 5 is a case in point: although A, B, and C are identifiable as exogenous and D and E as endogenous, the regularities encoded in that mosaic can be causally interpreted such that D is a cause of E, as in the minimal theory (13), or such that E is a cause of D, as in theories (14) and (15). The residual symmetry in Table 5 stems from the tight interdependence of the minimally sufficient conditions for D and E. Every such condition for D contains b or c, whereas every corresponding condition for E comprises B or C. Consequently, the minimal theories implied by Table 5 can be converted into one another by logically recombining the values of D and E with those of B and C. Moreover, whenever there is a change in a minimally sufficient condition for D, there is also a change in a minimally sufficient condition for E, and vice versa. Or, to use interventionist terminology: the causal (sub-)structures behind D and E are *not modular*, meaning it is not possible to manipulate one of those structures without simultaneously manipulating the other (Woodward, 2003, ch. 7.4).

To break the symmetries in the mosaic of Table 5, modularity must be restored, which requires expanding the mosaic with additional factors. One way of doing so, which bears a structural resemblance to interventionist approaches without presupposing interventionist assumptions or notions, is to add at least one further exogenous factor for each endogenous factor. In the context of Table 5, this means expanding the mosaic so that it contains at least one minimally sufficient condition for each of D and E, call them Y_D and Y_E, that appear in a minimal theory Ψ entailed by the expanded mosaic and whose factors appear only in one disjunct of the RDN-biconditionals for D and E, respectively, and nowhere else in Ψ. We refer to these modularity-restoring conditions as *specific disjuncts*.

To provide a more precise definition of the notion of a specific disjunct, we first have to slightly revise our definition of an RDN-biconditional. As shown in the previous section, requiring the disjunctions in RDN-biconditionals to contain at least two disjuncts does not provide a general solution to the directionality problem. Hence, we drop this requirement:

RDN-biconditional: A biconditional $\Phi \leftrightarrow Z$ entailed by a mosaic $\mathcal{M}$ is an RDN-biconditional if, and only if, according to $\mathcal{M}$, Φ is a minimally necessary disjunction of minimally sufficient conjunctions in DNF and Z is a single factor value.

Under this revised definition, an expression like $A \leftrightarrow Z$, if entailed by a mosaic, now counts as an RDN-biconditional, even though the regularities it encodes are entirely symmetric. This will not pose a problem for our account later on because our ultimate directionality criterion will be built on UMS, which catches symmetric RDN-biconditionals: the Humean mosaic that corresponds to $A \leftrightarrow Z$ (comprising exactly two rows: (A=1, Z=1); (A=0, Z=0)) has two MaxInd sets such that the complement does not change in isolation, namely {A} and {Z}, and thus violates UMS.

We can now define the notion of a specific disjunct as follows:

Specific disjunct: A conjunction Y_Z appearing as a disjunct in a disjunction Φ of an RDN-biconditional $\Phi \leftrightarrow Z$, which in turn is contained in a minimal theory Ψ, is a specific disjunct if, and only if, no factor in Y_Z appears anywhere else in Ψ.

A specific disjunct Y_Z is a minimally sufficient condition of Z composed of exogenous factors that can fix Z independently of the rest of the superordinate structure: if Y_Z is realized, Z is realized as well, regardless of how the remaining causal structure changes. In other words, Y_Z generates a context in which the remaining structure Ψ^{-}, obtained from Ψ by removing $\Phi \leftrightarrow Z$, can vary independently of $\Phi \leftrightarrow Z$. It follows that if a mosaic entails a minimal theory in which every effect has a specific disjunct, every effect has a context in which it can vary independently of the RDN-biconditionals of all other effects, namely the context in which all other effects are fixed by their specific disjuncts. This, in turn, means that the superordinate causal structure is modular.

Let us make this concrete by expanding the mosaic in Table 5 with two further exogenous factors I_D and I_E, yielding a total of five exogenous factors {A, B, C, I_D, I_E}, in addition to the two endogenous factors {D, E}. Also, let us assume that I_D is sufficient for D, that I_E is sufficient for E, and that the dependencies expressed in the minimal theory (13) obtain. If we list all $2^5 = 32$ logically possible configurations of the exogenous factors and assign values to D and E in accordance with these assumed dependencies, we get the mosaic in Table 7. Table 5 is contained in the last eight rows of Table 7, where both I_D and I_E take the value 0.

#	A	B	C	I_D	D	I_E	E
σ_1	1	1	1	1	1	1	1
σ_2	0	1	1	1	1	1	1
σ_3	1	0	1	1	1	1	1
σ_4	0	0	1	1	1	1	1
σ_5	1	1	0	1	1	1	1
σ_6	0	1	0	1	1	1	1
σ_7	1	0	0	1	1	1	1
σ_8	0	0	0	1	1	1	1
σ_9	1	0	1	0	1	1	1
σ_{10}	1	0	0	0	1	1	1
σ_{11}	0	0	0	0	1	1	1
σ_{12}	1	1	1	0	0	1	1
σ_{13}	0	1	1	0	0	1	1
σ_{14}	0	0	1	0	0	1	1
σ_{15}	1	1	0	0	0	1	1
σ_{16}	0	1	0	0	0	1	1
σ_{17}	1	1	1	1	1	0	1
σ_{18}	0	1	1	1	1	0	1
σ_{19}	1	0	1	1	1	0	1
σ_{20}	0	0	1	1	1	0	1
σ_{21}	1	1	0	1	1	0	1
σ_{22}	0	1	0	1	1	0	1
σ_{23}	1	0	0	1	1	0	0
σ_{24}	0	0	0	1	1	0	0
σ_{25}	1	0	1	0	1	0	1
σ_{26}	1	1	1	0	0	0	1
σ_{27}	0	1	1	0	0	0	1
σ_{28}	1	1	0	0	0	0	1
σ_{29}	0	1	0	0	0	0	1
σ_{30}	1	0	0	0	1	0	0
σ_{31}	0	0	0	0	1	0	0
σ_{32}	0	0	1	0	0	0	0

$$(A{\ast}b \;+\; b{\ast}c \;+\; I_D \;\leftrightarrow\; D){\ast}(B \;+\; C{\ast}D \;+\; I_E \;\leftrightarrow\; E) \qquad (13^{\ast})$$

Table 7: A Humean mosaic that results from integrating specific disjuncts into (13). It contains Table 5 in rows σ_{25} to σ_{32}, highlighted with grey shading; and it entails the minimal theory (13*).

That is, Table 5 can be thought of as obtained against a background in which I_D and I_E—two latent causal switches of D and E—are constantly absent.

Table 7 is the Humean mosaic that corresponds to the minimal theory (13*), which results from (13) by adding a specific disjunct I_D to the RDN-biconditional for D and a specific disjunct I_E to the RDN-biconditional for E. By contrast, there is no way to integrate I_D and I_E into minimal theories (14) and (15) such that the expanded theories remain compatible with Table 7. According to both (14) and (15), $E{\ast}b$ is sufficient for D, but D is not always present in Table 7 when the newly introduced switch I_E causes E in the absence of B. For example, configuration σ_{14} features $E{\ast}b$ together with d. More generally, the new possibility of fixing E independently of all other factors in the mosaic shows that E makes no difference to D and, hence, is not causally relevant for D, contrary to what is stipulated by (14) and (15). Conversely, none of the configurations in Table 7 in which D is independently fixed by the new switch I_D violate (13): in all of them it holds that whenever $C{*}D$ is given, so is E. D continues to make a difference to E.

To obtain Table 7, we introduced I_D and I_E into the mosaic of Table 5 in accordance with the minimal theory (13). Had we generated the expanded mosaic by adhering to the dependencies expressed in theories (14) or (15), it would not have entailed (13*) but corresponding expansions of (14) and (15):

$$(E{*}b \;+\; b{*}c \;+\; I_D \;\leftrightarrow\; D){*}(B \;+\; A{*}C \;+\; I_E \;\leftrightarrow\; E) \tag{14*}$$

$$(E{*}b \;+\; c{*}e \;+\; I_D \;\leftrightarrow\; D){*}(B \;+\; A{*}C \;+\; I_E \;\leftrightarrow\; E) \tag{15*}$$

Integrating specific disjuncts for all effects into a mosaic satisfying UMS is a generalizable strategy for breaking the remaining symmetries. The reason is that, once UMS has fixed a unique exogenous/endogenous partition and specific disjuncts restore modularity for all effects, opposed orientations among endogenous factors cannot survive in parallel. To show this, we now abstract from the concrete example. Consider a mosaic $\mathcal{M}$ over a factor set $\mathbf{F}$ that entails a clear separation between exogenous factors $\mathbf{F}_{\text{exo}}$ and endogenous factors $\mathbf{F}_{\text{end}}$. Suppose further that for two factors $Z_1, Z_2 \in \mathbf{F}_{\text{end}}$, $\mathcal{M}$ entails two minimal theories:

Ψ_1, containing the RDN-biconditionals

$$\Phi_1^{[Z_2]} \leftrightarrow Z_1 \quad \text{and} \quad \Phi_2^{\overline{[Z_1]}} \leftrightarrow Z_2,$$

where Z_2 appears in $\Phi_1^{[Z_2]}$ and Z_1 does *not* appear in $\Phi_2^{\overline{[Z_1]}}$, and

Ψ_2, containing the inverted RDN-biconditionals

$$\Phi_3^{\overline{[Z_2]}} \leftrightarrow Z_1 \quad \text{and} \quad \Phi_4^{[Z_1]} \leftrightarrow Z_2.$$

Thus, the mosaic $\mathcal{M}$ leaves it indeterminate whether Z_1 is causally relevant for Z_2 or vice versa. Still, we assume that, at bottom, there is one determinate direction of causation: the underlying causal structure is acyclic and is faithfully represented by some minimal theory that $\mathcal{M}$ entails. This true minimal theory may coincide with one of Ψ_1 and Ψ_2, or it may be another of the minimal theories $\mathcal{M}$ entails; in either case, the relation between Z_1 and Z_2 has a definite orientation.

To identify that direction, we expand $\mathcal{M}$ to a mosaic $\mathcal{M}^*$. For each effect in $\{Z_1, Z_2, \ldots, Z_m\}$, we add a new sufficient condition I_{Z_i} such that the corresponding factor I_{Z_i} does not appear in a minimally sufficient condition of any other effect Z_j, $j \neq i$. The new MaxInd set is $\mathbf{F}^*_{\text{exo}} = \mathbf{F}_{\text{exo}} \cup \{\mathrm{I}_{Z_1}, \ldots, \mathrm{I}_{Z_m}\}$, and $\mathcal{M}^*$ has $2^{|\mathbf{F}^*_{\text{exo}}|}$ configurations. Any minimal theory entailed by $\mathcal{M}^*$ must incorporate these newly introduced sufficient conditions as specific disjuncts. Introducing such disjuncts into Ψ_1 or Ψ_2 amounts to replacing all component RDN-biconditionals $\Phi_i \leftrightarrow Z_i$ with $\Phi_i + I_{Z_i} \leftrightarrow Z_i$, resulting in expanded theories Ψ^*_1 and Ψ^*_2, which contain the following RDN-biconditionals:

$$\Psi^*_1: \quad \Phi_1^{[Z_2]} + I_{Z_1} \leftrightarrow Z_1 \qquad (\mathrm{RDN}^1_{Z_1})$$

$$\Phi_2^{\overline{[Z_1]}} + I_{Z_2} \leftrightarrow Z_2 \qquad (\mathrm{RDN}^1_{Z_2})$$

$$\Psi^*_2: \quad \Phi_3^{\overline{[Z_2]}} + I_{Z_1} \leftrightarrow Z_1 \qquad (\mathrm{RDN}^2_{Z_1})$$

$$\Phi_4^{[Z_1]} + I_{Z_2} \leftrightarrow Z_2 \qquad (\mathrm{RDN}^2_{Z_2})$$

Theorem 1. *If $\mathcal{M}^*$ satisfies UMS and is an expansion of $\mathcal{M}$ that introduces a specific disjunct for every effect, then $\mathcal{M}^*$ cannot entail both Ψ^*_1 and Ψ^*_2.*

Proof. Since the underlying causal structure is acyclic, it is faithfully represented by a permanently redundancy-free minimal theory entailed by $\mathcal{M}$. By the permanence condition built into the regularity-theoretic definition of causal relevance, its Boolean determination relations are preserved under expansions; hence, $\mathcal{M}^*$ entails a corresponding expansion in which the specific disjuncts add further sufficient conditions without rendering any original determination relation redundant. Thus, the true orientation between Z_1 and Z_2 survives in $\mathcal{M}^*$; what remains to be shown is that $\mathcal{M}^*$ cannot entail both Ψ^*_1 and Ψ^*_2.

To this end, we first focus on the RDN-biconditionals for Z_1, $(\mathrm{RDN}^1_{Z_1})$ and $(\mathrm{RDN}^2_{Z_1})$. $(\mathrm{RDN}^1_{Z_1})$ expresses Z_1 in terms of the minimized Boolean function $\Phi_1^{[Z_2]}$, which contains Z_2 (non-trivially). For $\mathcal{M}^*$ to entail $(\mathrm{RDN}^1_{Z_1})$, it must therefore contain rows in which the factors in $\Phi_1^{[Z_2]}$ other than Z_2 take constant values and a change in Z_2 changes the truth value of $\Phi_1^{[Z_2]}$. The truth value of $\Phi_1^{[Z_2]}$ can thus be changed directly by merely changing the value of Z_2. By contrast, ($\text{RDN}^2_{Z_1}$) expresses Z_1 in terms of the minimized Boolean function $\Phi_3^{\overline{[Z_2]}}$, which does not contain Z_2. A change in Z_2 therefore cannot directly change the truth value of $\Phi_3^{\overline{[Z_2]}}$. It could do so only indirectly, by inducing a change in the value of some other endogenous factor, say Z_k, contained in $\Phi_3^{\overline{[Z_2]}}$.

Now, if $\mathcal{M}^*$ were to entail both Ψ_1^* and Ψ_2^*, it would have to entail both ($\text{RDN}^1_{Z_1}$) and ($\text{RDN}^2_{Z_1}$). Since these biconditionals share the same right-hand side, their left-hand sides would then have to take identical truth values in every row of $\mathcal{M}^*$. But modularity rules this out. Because every effect Z_i can be fixed independently by its specific disjunct I_{Z_i} without affecting the causal structures behind the other effects, I_{Z_2} can change the value of Z_2 while all other endogenous factors are held fixed by their own specific disjuncts. Some of these changes directly affect the truth value of $\Phi_1^{[Z_2]}$, but none affect the truth value of $\Phi_3^{\overline{[Z_2]}}$. Even if $\Phi_3^{\overline{[Z_2]}}$ contains some endogenous factor Z_k that depends on Z_2, UMS ensures that there will be rows in $\mathcal{M}^*$ in which Z_k is fixed via I_{Z_k} while the value of Z_2 is changed by I_{Z_2}. In those rows, the truth value of $\Phi_1^{[Z_2]}$ changes whereas the truth value of $\Phi_3^{\overline{[Z_2]}}$ does not. Hence, the left-hand sides of ($\text{RDN}^1_{Z_1}$) and ($\text{RDN}^2_{Z_1}$) cannot agree in all rows of $\mathcal{M}^*$, and so ($\text{RDN}^1_{Z_1}$) and ($\text{RDN}^2_{Z_1}$) cannot both be entailed by $\mathcal{M}^*$.

The same reasoning applies to ($\text{RDN}^1_{Z_2}$) and ($\text{RDN}^2_{Z_2}$). Their left-hand sides, too, differ in that one contains the relevant endogenous factor and the other does not. Accordingly, they also cannot take identical truth values in all rows of $\mathcal{M}^*$. It follows that $\mathcal{M}^*$ cannot entail both Ψ_1^* and Ψ_2^*. Therefore, once specific disjuncts restore modularity in a mosaic satisfying UMS, the remaining symmetry between Z_1 and Z_2 is broken. $\square$

The theorem just established yields the following sufficient directionality criterion for regularity theories:

Regularity Directionality (RD): All causal relations obtaining in a Humean mosaic $\mathcal{M}$ over a factor set **F** can be unambiguously oriented if the following conditions are met:

(1) **Unique MaxInd Set Condition (UMS):** $\mathcal{M}$ has a unique MaxInd set such that the factors in the complement of that set do not change their values in isolation.

(2) **Specific Disjunct Condition (SDC):** $\mathcal{M}$ entails a minimal theory Ψ containing a specific disjunct Y_{Z_i} for every effect Z_i.

Two points deserve emphasis. First, RD guarantees unambiguous orientability but not uniqueness of causal structure in the stronger sense that a mosaic satisfying RD entails exactly one minimal theory. RD is compatible with a plurality of minimal theories being entailed by a mosaic, as long as these theories do not yield opposite causal orientations. Second, RD provides a *sufficient* condition for unambiguous orientability, not a *necessary* one. It may be possible to identify the direction of causation from a Humean mosaic with only some specific disjuncts present, or even without any at all. The mosaic in Table 6 is a case in point. To effectively counter Lewis' (1973) criticism that the direction of causation cannot always be inferred from regularities alone, a condition is needed that *guarantees* unambiguous orientability for *all* acyclic causal structures. This is what RD delivers.

To get an intuitive grip on how the introduction of specific disjuncts can break symmetries, let us return to the simple mosaic in Table 3 and its interpretation in terms of a staircase light (C) with two switches (A, B). That mosaic is fully symmetric: not only can the light be expressed as a function of the two switches, but each switch can also be expressed as a function of the light and the other switch. Suppose now that we expand this structure so that the light can also be turned on by an independent circuit D, for example, a backup battery or a motion sensor. The mosaic resulting from this expansion is in Table 8. The backup circuit being active, D, constitutes a specific disjunct for the light, since D is sufficient for C regardless of how switches A and B are set, and it is completely independent of the switches

#	A	B	D	C
σ_1	1	1	1	1
σ_2	1	0	1	1
σ_3	0	1	1	1
σ_4	0	0	1	1
σ_5	1	1	0	1
σ_6	0	0	0	1
σ_7	0	1	0	0
σ_8	1	0	0	0

$$A{*}B \; + \; a{*}b \; + \; D \;\leftrightarrow\; C \qquad (9^*)$$

Table 8: A Humean mosaic that results from integrating a specific disjunct into (9). It contains Table 3 in rows σ_5 to σ_8, highlighted with grey shading; and it entails the unique minimal theory (9*).

and their relation to the light. Its introduction breaks the original symmetry because none of the switches can be expressed as a function of the other switch and the light in the expanded mosaic, as the light being on no longer allows for inferring back to the switch positions. C is the only factor expressible as a function of the other factors in Table 8. Correspondingly, that mosaic entails a unique causal orientation, expressed in the minimal theory (9^*).

As noted above, expanding mosaics to include specific disjuncts is only one of several conceivable strategies to ensure unambiguous orientability. Nevertheless, it is a strategy that has a longstanding tradition in another theoretical framework: interventionism. Interventionist theories of causation secure modularity by assuming that there exist intervention variables outside of a modeled system for all effects within that system, and they hold that the existence of possible interventions is a defining feature of causation (Woodward, 2003, 59). Specific disjuncts are regularity-theoretic analogues of intervention variables, with the significant difference that they are defined in purely functional terms, whereas intervention variables are defined in causal terms (Woodward, 2003, 98).

RD can be integrated into a regularity theory by following interventionism in assuming that the causal complexity of real-world systems is sufficiently high to ensure the existence of surgical sufficient conditions for all effects, whether through experimental manipulation, natural experiments, or engineered interventions. This *sufficient-complexity assumption* entails that Humean mosaics of real-world systems can, in principle, always be expanded by specific disjuncts. Consequently, minimal theories that do not contain specific disjuncts in all RDN-biconditionals do not represent the whole system (even if the causal relations they entail are unambiguously orientable); we call them *unsaturated* theories. In contrast, minimal theories that explicitly feature specific disjuncts in all RDN-biconditionals are termed *saturated*. Against that background, (CR) can be sharpened by requiring that causal relevance of A for B be witnessed by a permanently redundancy-free minimal theory in which A figures in an RDN-biconditional for B, and that this orientation be non-reversible through saturation, meaning that no saturated minimal theory entailed by an expansion of the mosaic reverses it. As shown above, this non-reversal is secured for mosaics satisfying (RD). This guarantees that every non-cyclic causal relation receives an unambiguous orientation without reliance on temporal precedence.

The sufficient-complexity assumption will be challenged by critics of interventionism and causal modularity (e.g. Cartwright, 2002). While we do not have the space to properly enter that debate here, we want to offer some argumentative support for the assumption. On the one hand, it can be grounded in the success of scientific practice. Controlled experimentation critically relies on our ability to manipulate one factor while holding others fixed—that is, it relies on the modularity of the studied systems. Hence, the success of experimental science suggests that the world is indeed causally modular. In fields like econometrics and epidemiology, which deal with complex systems where direct manipulation is often impossible, researchers rely on so-called *instrumental variables*, which amount to non-experimental analogues to specific disjuncts (Angrist and Pischke, 2009). The widespread success in discovering instrumental variables substantiates that specific disjuncts are not just theoretical idealizations but are real features of complex systems. On the other hand, the sufficient-complexity assumption receives support from the causal openness of real-world systems, which are embedded in a vast web of potential influencing factors. For any given effect within a local model, there exists—outside of the model—a near-infinite pool of exogenous factors that are candidates for figuring in specific disjuncts.

Ultimately, the sufficient-complexity assumption serves as a regulative ideal. Building a theory of causation on it entails that causation is only well-defined for complex systems. This shifts the burden of analysis. Previously, if a regularity theory failed to identify a direction, it was viewed as a failure of the definition of causation. This assumption reframes the issue: if we cannot identify direction, it is because we have not yet found the saturated minimal theory. The failure is epistemic, not conceptual.

5 Conclusion

It has long been a central criticism of regularity theories of causation that they are incapable of distinguishing causes from effects without appealing to an external asymmetry, such as temporal precedence. This problem, famously articulated by Lewis (1973), arises because the Boolean determination relations of sufficiency and necessity, which constitute the analytical core of the theory, can be symmetric even when the underlying causal structure is not. We have demonstrated that the standard response of requiring effects to be deterministic colliders is inadequate, as Humean mosaics can satisfy this collider criterion yet entail minimal theories with opposed causal orientations.

To solve this directionality problem, we introduced the Regularity Directionality (RD) criterion. RD provides a sufficient condition for unambiguous orientability. Its first component, the Unique MaxInd Set Condition (UMS), fixes a unique partition of the analyzed factors into exogenous and endogenous ones, while its second component, the Specific Disjunct Condition (SDC), removes the remaining symmetries among endogenous factors by requiring a specific disjunct for every effect. RD does not imply uniqueness of causal structure in the stronger sense that a mosaic must be modelable by exactly one minimal theory; rather, it excludes opposed causal orientations. We propose to integrate RD into the regularity framework via a sufficient-complexity assumption: real-world systems are sufficiently complex to furnish specific disjuncts for all effects. By amending the definition of causal relevance (CR) with the requirement that a causal orientation be non-reversible in any saturated minimal theory entailed by an expansion of the mosaic, it can be ensured that every acyclic causal relation receives an unambiguous orientation.

This paper has focused on solving the directionality problem for type causation. A natural next step is to investigate how our solution can be extended to orient causal relations at the token level. Moreover, that RD assigns a crucial analytical role to specific disjuncts indicates that there are more points of contact between modern regularity theories and interventionist theories of causation than previously recognized. Further exploring synergies between these frameworks is a promising avenue for future research. Finally, the theoretical solution offered here invites practical application. It has been developed for complete and noise-free Humean mosaics; real data tend not to meet these conditions. Hence, operationalizing RD requires adapting the criterion to incomplete and noisy mosaics, for example, by relaxing the completeness built into UMS and tolerating approximate rather than exact specific disjuncts. Moreover, the operationalized RD criterion should be implemented in computational methods that detect MaxInd sets and specific disjuncts in empirical data. This would transform RD from a philosophical principle into an instrument for causal learning.

References

Andreas, H. and M. Guenther (2021). Regularity and inferential theories of causation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.). Metaphysics Research Lab, Stanford University.

Andreas, H. and M. Günther (2024). A regularity theory of causation. *Pacific Philosophical Quarterly 105*(1), 2–32.

Angrist, J. D. and J.-S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.

Baumgartner, M. (2008a). Regularity theories reassessed. *Philosophia 36*, 327–354.

Baumgartner, M. (2008b). The causal chain problem. *Erkenntnis 69*, 201–226.

Baumgartner, M. (2013). A regularity theoretic approach to actual causation. *Erkenntnis 78*, 85–109.

Baumgartner, M. and C. Falk (2023). Boolean difference-making: A modern regularity theory of causation. *The British Journal for the Philosophy of Science 74*(1), 171–197.

Beebee, H. (2000). The non-governing conception of laws of nature. *Philosophy and Phenomenological Research 61*(3), 571.

Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two: Comments on Hausman and Woodward. *British Journal for the Philosophy of Science 53*, 411–453.

Csikszentmihalyi, M. (1975). *Beyond Boredom and Anxiety*. San Francisco: Jossey-Bass Publishers.

De Souter, L. (2025). A reductive INUS theory of causation. Unpublished manuscript, June 2025. Available at `https://luna-de-souter.github.io/assets/Reductive_INUS_theory.pdf`.

Dummett, M. (1964). Bringing about the past. *The Philosophical Review 73*(3), 338–359.

Graßhoff, G. and M. May (2001). Causal regularities. In W. Spohn, M. Ledwig, and M. Esfeld (Eds.), *Current Issues in Causation*, pp. 85–114. Paderborn: Mentis.

Hume, D. (1999 (1748)). *An Enquiry Concerning Human Understanding*. Oxford: Oxford University Press.

Kim, J. (1971). Causes and events: Mackie on causation. *Journal of Philosophy 68*, 426–441.

Lewis, D. (1973). Causation. *Journal of Philosophy 70*, 556–567.

Lewis, D. (1986). *Philosophical Papers II*. Oxford: Oxford University Press.

Mackie, J. L. (1974). *The Cement of the Universe. A Study of Causation*. Oxford: Clarendon Press.

Mackie, J. L. (1993 (1965)). Causes and conditions. In E. Sosa and M. Tooley (Eds.), *Causation*, pp. 33–55. Oxford: Oxford University Press.

Mano, M. M. and M. D. Ciletti (2012). *Digital Design* (5th ed.). Upper Saddle River, NJ: Prentice Hall.

Mill, J. S. (1843). *A System of Logic*. London: John W. Parker.

Psillos, S. (2010). Regularity theories. In H. Beebee, C. Hitchcock, and P. Menzies (Eds.), *The Oxford Handbook of Causation*, pp. 131–157. Oxford University Press.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Spirtes, P., C. Glymour, and R. Scheines (2000). *Causation, Prediction, and Search* (2 ed.). Cambridge: MIT Press.

Taylor, R. (1964). Can a cause precede its effect? *The Monist 48*(1), 136–142.

Weinberger, N., P. Williams, and J. Woodward (2024). The worldly infrastructure of causation. *The British Journal for the Philosophy of Science*. doi: 10.1086/730698

Woodward, J. (2003). *Making Things Happen. A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wright, R. W. (2011). The NESS account of natural causation: A response to criticisms. In R. Goldberg (Ed.), *Perspectives On Causation*, pp. Chapter 14. Hart Publishing.

Zhang, J. and K. Zhang (2023). A new minimality condition for Boolean accounts of causal regularities. *Erkenntnis 90*(1), 67–86.