Causality as a Theoretical Concept: Explanatory Warrant and Empirical Content of the Theory of Causal Nets*

Gerhard Schurz and Alexander Gebharter

Abstract: We start this paper by arguing that causality should, in analogy with force in Newtonian physics, be understood as a theoretical concept that is not explicated by a single definition, but by the axioms of a theory. Such an understanding of causality implicitly underlies the well-known theory of causal (Bayes) nets (TCN) and has been explicitly promoted by Glymour (2004). In this paper we investigate the explanatory warrant and empirical content of TCN. We sketch how the assumption of directed cause-effect relations can be philosophically justified by an inference to the best explanation. We then ask whether the explanations provided by TCN are merely post-facto or have independently testable empirical content. To answer this question we develop a fine-grained axiomatization of TCN, including a distinction of different kinds of faithfulness. A number of theorems show that although the core axioms of TCN are empirically empty, extended versions of TCN have successively increasing empirical content.

Keywords: screening off, linking up, axioms for causal nets, inference to the best explanation, empirical content

1. Introduction and formal preliminaries

Cognitive biologists tell us that causal reasoning is an evolutionary highly successful characteristic of homo sapiens (Tomasello 1999). This diagnosis stands in stark contrast to the perennial difficulties philosophers had when trying to *justify* causality as something ontologically real. According to Hume's fundamental skeptical challenge, the classification of correlated events into causes and effects doesn't correspond to anything real "out there in the world" – only the

* This is a draft paper. The final version of this paper is published under the following bibliographical data: Schurz, G., & Gebharter, A. (2016). Causality as a theoretical concept: Explanatory warrant and empirical content of the theory of causal nets. *Synthese, 193*(4), 1073-1103. doi:10.1007/s11229-014-0630-z. The final publication is available at http://link.springer.com/.

correlations are real, while causality is a mere habit of our cognitive system. Many philosophers have tried to answer Hume's challenge, but none of the current accounts of causality has become commonly accepted.[1]

Most philosophical approaches have attempted to explicate the concept of causality by means of definitions. According to Glymour (2004), this is a shortcoming. He compares definitional ("Socratic") with axiomatic ("Euclidean") approaches to causality; the latter ones being exemplified in the theory of causal (Bayes) nets (TCN). Glymour argues that axiomatic accounts are more fruitful than definitional ones. We agree with Glymour and take his (2004) paper as our starting point. We suggest that causality should, in analogy to the concept of Newtonian *force*, be understood as a *theoretical concept*. It is a characteristic property of theoretical concepts that their meaning cannot be fully specified by a single definition, but only by the joint effect of the core axioms of a *theory*. In case of *force*, this theory is Newtonian mechanics. We argue that in case of causality, this theory should be TCN.

The paper is structured as follows. In sec. 2 we propose an answer to Hume's challenge by providing a philosophical justification of causality (as axiomatized by TCN) as something ontologically real by an inference to the best explanation (IBE) for two statistical phenomena: screening off and linking up. In sec. 3 we investigate the question of whether TCN and various extended TCN-versions have independently testable empirical content, i.e. exclude some logically possible probability distributions. The core principles of TCN (causal Markov and minimality) turn out to be empirically empty. TCN-versions which additionally assume faithfulness or external noise produce empirical content, which – on pains of avoiding falsification – is only probabilistic in nature. More empirical content (including deductive content) is gained by assuming temporal forward-directedness.

---

[1] For renewals of Hume's challenge cf. Psillos (2009) and Norton (2009). For supporters of causality as something real see Beebee et al. (2009, parts II and III).

TCN has been developed by SGS (= Spirtes, Glymour, and Scheines) (2000) and Pearl (1988, 2009), with forerunners such as Wright (1921), Reichenbach (1956), Blalock (1961), and Suppes (1970). Our paper intends to contribute to this theory with respect to the following points:

(1.) Our understanding of causality as a theoretical concept axiomatized by TCN and justified by its explanatory success is implicitly held by many proponents of TCN. In sec. 2 we go a step further by arguing that directed cause-effect relations as axiomatized by TCN are the *best* explanation of two probabilistic phenomena: screening off and linking up. We consider alternative explanations and highlight their problems and disadvantages.

(2.) In sec. 3 we apply methods, originally developed within philosophy of science for investigating the empirical (or non-theoretical) content of scientific theories, to the theory of causal nets. We feel that this move constitutes a new and fruitful enterprise within the TCN research program.

(3.) From (2.) we obtain a variety of results. Some of these results are technically known, but their philosophical consequences are new or deserve reconsideration (th. 1, 3, 4, and our distinction between "stable" vs. "unstable" kinds of unfaithfulness), while other results are also technically new (th. 2, 5, 6).

In the remainder of this section we introduce probabilistic (in)dependence, which constitute the central empirical (or non-theoretical) concepts of TCN. We understand (Humean) regularities in a broad *probabilistic* sense as probabilistic *dependencies* (or correlations). We interpret probabilities of repeatable events as their dispositions to occur with corresponding limiting frequencies – this interpretation is important for our attempt to justify causal connections by their power to explain probabilistic dependencies as objective features of the world (as opposed to subjective features of beliefs). An intended consequence of this view is that causal claims involving singu-

lar events have to be backed up by probabilistic regularities.

Our account makes use of mathematical variables. A variable is a function X: D→Val(X) from a domain D of individuals to its value space Val(X) = $\{x_1,x_2,\ldots\}$, which is a family of properties or a set of numbers. If X denotes color, for example, then Val(X) = {red,green,…} and X assigns a color X(d) to every individual d∈D. That d has color green may be expressed by "X(d) = x_2", where "x_2" stands for green. Note that properties or event-types are not variables, but values of variables. We also admit that D consists of n-tuples of individuals, e.g. individuals at certain time-points. Simple dichotomic property-pairs are represented by binary variables X_F with value space {F,¬F} (e.g., {red,not-red}). We make use of the following notational conventions:

- X,Y,… are variables and **U**,**V**,**W** in bold letters are sequences of variables.
- Lower-case letters "x" (or "x_i") stand for values of X; lower-case "**u**" (or "$\mathbf{u}_i$") for sequences of values of variables in **U**, i.e. $\mathbf{u}\in Val(\mathbf{U}) = Val(X_1)\times\ldots\times Val(X_n)$ if $\mathbf{U} = (X_1,\ldots,X_n)$.
- $P(X_1,\ldots X_n)$ is a (statistical) probability distribution over a suitable algebra AL over the space of values, i.e. $P: AL(Val(X_1)\times\ldots\times Val(X_n))\rightarrow[0,1]$.
- "P(x)" abbreviates "P({x})" and stands for "P(X(α) = x)"; so P(x) is the probability that X takes value x in the underlying domain D.[2]
- "P(x∈S)" abbreviates "P(X(α)∈S)" (where $S \subseteq Val(X)$), i.e. the probability that the value of X (in domain D) lies in the value-range S.

---

[2] $P(X_1)$ is defined from $P(X_1,\ldots,X_n)$ by the usual projection postulate. α in P(X(α) = x) is an individual variable that is *bound* by the probability operator P (we use the letter "α" because "x" is reserved for X-values). In the *statistical* interpetation, P(x) is the limiting frequency of result x in an infinite sequence of random drawings of individuals α from D. This covers also the *generic propensity* interpretation, in which one interprets P(x) as the limiting frequency of result x in an infinite sequence of performances of a random experiment; here D consists of the individual performances of the experiment. In the *single propensity* interpretation, in contrast, P is attached to individual events (such as *this* throwing of *this* coin); here P is assumed as a primitively given function over $AL(\Pi_{1\le i\le n}Val(X_i))$.

- "$P(\neg x)$" abbreviates "$P(X(\alpha) \neq x)$", "$P(x,y)$" abbreviates "$P(X(\alpha) = x \wedge Y(\alpha) = y)$", and
- "$P(x|y)$" abbreviates "$P(X(\alpha) = x|Y(\alpha) = y)$", i.e. the conditional probability of x given y, provided $P(y) > 0$.[3]

Two variables X,Y are said to be probabilistically dependent (DEP(X,Y)) iff at least *some* of their values are dependent; they are probabilistically independent (INDEP(X,Y)) iff *all* of their values are independent. Thus, probabilistic (in)dependence between variables can be defined by any one of the following equivalent formulations (a)-(c):

(1) DEP(X,Y) *iff*

(a) $\exists x,y: P(x|y) \neq P(x)$ and $P(y) > 0$, or

(b) $\exists x,y: P(y|x) \neq P(y)$ and $P(x) > 0$, or

(c) $\exists x,y: P(x,y) \neq P(x) \cdot P(y)$.

INDEP(X,Y) *iff* not DEP(X,Y), i.e. *iff* $\forall x,y: P(x|y) = P(x)$ or $P(y) = 0$ (with equivalent formulations similar to the formulations above).

The equivalence of (a) with (b) makes clear that probabilistic dependencies are always *symmetric*. *Conditional* (in)dependence between X and Y given variables $Z_1,\ldots,Z_n$ is defined as follows:

(2) $DEP(X,Y|Z_1,\ldots,Z_n)$ *iff* $\exists x,y,z_1,\ldots,z_n: P(x|y,z_1,\ldots,z_n) \neq P(x|z_1,\ldots,z_n)$ and $P(y,z_1,\ldots,z_n) > 0$ (with equivalent formulations similar to the formulations in (1)).

$INDEP(X,Y|Z_1,\ldots,Z_n)$ *iff* not $DEP(X,Y|Z_1,\ldots,Z_n)$.

---

[3] $P(y) > 0$ is required because $P(x|y)$ is defined as $P(x,y)/P(y)$. If one wants to cover the case $P(y) = 0$, one may assume independently axiomatized conditional probabilities (cf. Pearl 2000, 11; Carnap 1971, 38f).

Unconditional (in)dependence (IN)DEP(X,Y) coincides with (IN)DEP(X,Y|∅), where ∅ is the empty set.

The definition of probabilistic dependence is generalized to sequences of variables **U**,**V**,**W** via the following definition: DEP(**U**,**V**|**W**) iff ∃**u**,**v**,**w**: P(**u**|**v**,**w**) ≠ P(**u**|**w**) and P(**v**,**w**) > 0.

DEP(X,Y) merely asserts the existence of dependencies between *some* values of X and Y. These dependencies can be positive or negative and of arbitrary form and degree. INDEP(X,Y), on the other hand, requires that *all* values of X and Y are independent.

Note that the notions of *positive* and *negative* probabilistic (in)dependence can, prima facie, only be defined for values of variables:

(3) POSDEP(x,y) *iff* P(x|y) > P(x); NEGDEP(x,y) *iff* P(x|y) < P(x);
DEP(x,y) *iff* POSDEP(x,y) or NEGDEP(x,y);
INDEP(x,y) *iff* ¬DEP(x,y).

However, we can define a notion of probabilistic dependence between variables if their values are ordered according to size: in that case we say that a *variable* Y is positively dependent on X *iff* an increased X-value leads to an increased mean or expectation value of Y.

2. Explanatory warrant of TCN: justifying causality by IBE

2.1 Causality as a theoretical concept: a comparison with Newtonian force

According to the findings of contemporary (post-positivistic) philosophy of science, scientific

theories contain theoretical concepts such as atom, force, etc.[4] Theoretical concepts are neither definable in terms of observable phenomena – they offer unified explanations of such phenomena in terms of hidden "deep structures" instead –, nor are they definable by a single theoretical principle or axiom. Rather, their semantic content is characterized by a theory, or at least by a *theory core* to which said concepts belong. Classical physics, for example, stipulates gravitational forces as unobservable causes of trajectories of physical bodies. The "meaning" of "gravitational force" is not determined by a single definition, but by the joint effect of the synthetic axioms of Newtonian mechanics which, when combined, entail a large variety of empirical consequences. We suggest that causality should, in precise analogy to force, be understood as a theoretical concept whose meaning can be explicated by TCN's core axioms. Thus, the empirical (or non-theoretical) part of TCN is the concept of a probability distribution over a set of variables, whose properties are to be explained by assuming theoretical cause-effect relations between these variables according to the principles of TCN.

In order to be empirically significant, theoretical concepts and the principles characterizing them must have the following two features:

(i) They offer unifying explanations of empirical phenomena which cannot be generated without them (explanatory warrant), and

(ii) they are not entirely ex-post, but generate empirical predictions by which they are independently testable (empirical content).

There is no guarantee that a theoretical concept refers to a really existing entity – purely *instrumental* interpretations of theoretical concepts as useful means for unifying empirical phenomena are always possible. But the more empirically successful a theory becomes, the more plausible it

[4] Cf. Carnap (1956), Lewis (1970), Sneed (1971), Balzer et al. (1987), Papineau (1996), French (2008).

is to assume that the theoretical concepts producing this success actually *do* refer to something real. The concept of force in Newtonian physics, for example, does have both features of empirical significance to an admirably large extent – presumably, no physicist (and only a few philosophers of science) would doubt that forces are real. In this paper we ask whether the concept of causality has the two ingredients of empirical significance, explanatory warrant and empirical content.

In this section (sec. 2) we investigate the explanatory warrant of TCN. The decisive question concerning TCN's explanatory warrant is: *What* does causality explain? The answer cannot be that every empirical regularity is explained by an underlying causal power. Of course, for every regularity one can postulate a corresponding causal connection that "explains" it *post facto*. But causal "explanations" of this sort would amount to a mere metaphysical duplication of empirical regularities that can neither achieve unification of regularities, nor generate novel predictions. Since they fail to meet (i) and (ii), Ockham's razor dictates to eliminate them.

Causality is also not needed to explain why observed regularities are inductively projectible, as some philosophers have suggested (cf. Fales 1990, ch. 4). The inductive projectibilty of regularities is already explained by assuming that they are backed up by *lawlike* connections (Armstrong 1983, part 1). Causality, however, goes *beyond* inductive projectibility or lawlikeness: regularities connecting the joint effects of a common cause, for instance, may be perfectly lawlike though obviously non-causal.

To withstand Hume's skeptical challenge one has to answer the question of why cause-effect relations are *needed at all*, instead of simply accepting lawlike regularities as primitive facts. Our answer is that cause-effect relations (as characterized by the core axioms of TCN) yield the best available explanation for two otherwise mysterious (in)stability properties of probabilistic regularities w.r.t. conditionalization: screening off and linking up.

2.2 Explaining screening off and linking up

Since screening off and linking up are the major explananda of causal relations, we have to characterize them in an empirical, i.e. purely probabilistic way, without presupposing causal notions.

(4) X and Y are screened off by Z *iff* DEP(X,Y) and INDEP(X,Y|Z).

Examples:

(4.1) Barometer reading (X) storm coming (Y) atmospheric pressure (Z)

(4.2) Young male (X) car accident (Y) car speed(Z)

The probabilistic dependence between X and Y disappears when conditionalizing on arbitrarily chosen but *fixed* values of a third variable Z. Condition (4) implies the probabilistic dependencies DEP(X,Z) and DEP(Y,Z). We assume the usual case that these dependencies are not themselves screened off by the respective third variable (Y and X, resp.). Moreover, we focus on *robust* (or faithful) cases of screening off in which the disappearance of the probabilistic X-Y dependence after conditionalization on Z is *stable* under small changes of the involved conditional probabilities (we shall see in sec. 3.2 that most cases of screening off are robust in this sense.)

Intuitively we interpret the correlations in (4.1) and (4.2) immediately as produced by causal relations: we believe that we "know" that screening off occurs because Z is a common cause in (4.1) and an intermediate cause in (4.2). In order to achieve a philosophical justification of causality we must free our mind from such prefabricated causal intuitions and assume for a moment that we only know the variables' probability distributions. If we do that, we are confronted with a riddle: *Why* does the X-Y correlation disappear when fixing Z's value?

The *best* available explanation of robust screening off phenomena – in fact, the only good explanation we can think of – is the following: only the two dependencies between X and Z and

between Z and Y reflect a direct "causal" connection between the respective variables,[5] while the dependence between X and Y results from these two causal connections and, thus, is *mediated* (or *transmitted*) by Z. This situation is depicted in fig. 1. Hence, if we consider subsets of individuals with different X-values, these individuals will have differently distributed Y-values only because they have differently distributed Z-values. So if we conditionalize on a subdomain of individuals with fixed Z-values, individuals with different X-values will no longer have differently distributed Y-values, i.e. the probabilistic dependence will no longer be transmitted from X to Y.

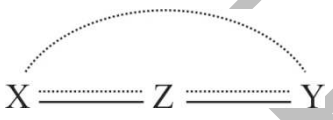


Fig. 1 Explanation of screening off by binary causal relations ("···" stands for a probabilistic dependence and "—" for a direct causal connection)

Note that explaining screening off only requires the assumption of an *undirected binary* dependence relation between variables; we call this dependence relation "causal" although no direction of causation is needed so far. Directed causation is, however, required for discriminating screening off from linking up. Before we come to this point, we show why prominent alternative attempts fail in explaining screening off.

First of all, *duplication accounts* cannot explain screening off. They come in two versions: (i) Humean-reductionist (causality is "nothing but" correlation) and (ii) naive-metaphysical (every correlation is "backed up" by a corresponding causal connection). Duplication accounts would postulate a direct causal connection between *every* two correlated variables X and Y, as shown in fig. 2(a). But then they cannot explain why Z screens off X from Y; this fact would remain mysterious. It is precisely the assumption that *not all* correlations correspond to direct causal connec-

[5] The claim that the causal connection between X and Z in fig. 1 is "direct" is relative to the set of variables {X,Y,Z}.

tions which explains screening off.

A second alternative explanation attempt would be the *blocking* account: Z can influence the causal connection between X and Y in such a way that some Z-values block this connection, as depicted in fig. 2(b). But this hypothesis cannot explain why the X-Y correlation vanishes when conditionalizing on *arbitrary* Z-values. It seems that the only explanation that works is the one given above: Z screens X off from Y because Z mediates X's dependence on Y.

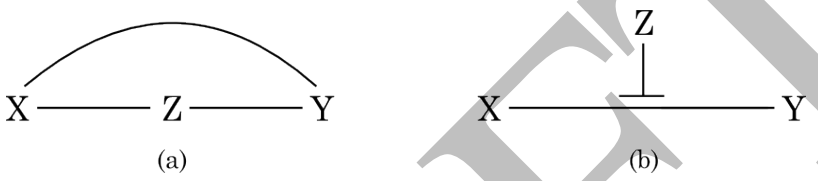


Fig. 2: (a) Duplication accounts cannot explain why DEP(X,Y) vanishes at all. (b) The blocking theory cannot explain why DEP(X,Y) vanishes when conditionalizing on arbitrary Z-values.

A final objection might point out that it is impossible even to *ask* for an explanation of screening off without already presupposing causal notions, because all explanations *are* causal. However, we don't understand the explanation of screening off in a causal sense (for otherwise we would end up in an infinite regress). In accordance with many philosophers of science, e.g. Friedman (1974) or Kitcher (1989), we assume that there is a non-causal sense of "explanation" consisting in unification and the generation of potential predictions.

Let us now turn to the question of how to explain linking up. Some sets of variables {X,Y,Z} have probability distributions that feature exactly the opposite (in)stability properties to screening off. We call this phenomenon "linking up" and define it again in a purely probabilistic (i.e. non-causal) way:

(5) X and Y are linked up by Z *iff* INDEP(X,Y) and DEP(X,Y|Z).

Example: Angle of the sun (X) length of a tower (Y) length of its shadow (Z)

Two independent variables X and Y become linked up by Z *iff* they become dependent after conditionalization on some values of Z. The position of the sun, for example, is not correlated with the height of a tower, but it becomes correlated if we conditionalize on the shadow's length. If the tower's shadow is long, for instance, we can infer that the solar altitude must be low if the tower is short.[6] As in screening off scenarios, (5) implies DEP(X,Z) and DEP(Y,Z). Again we focus on robust cases of linking up.

Let us once more put aside prefabricated causal intuitions. Then we face a second riddle: Why do two formerly independent variables X and Y become correlated when conditionalizing on certain Z-values? It is clear that undirected causal relations cannot explain *both* screening off and linking up. To explain linking up, Z must again act as a *mediator* between X and Y. So the structure of undirected causal relations in the linking up scenario must have the *same* form as in the screening off scenario depicted in fig. 1. But if the causal structure should be able to explain both screening off and linking up, it cannot have the same form in these two cases, because the two phenomena involve opposite probabilistic (in)stability effects.

The best available explanation for screening off *and* linking up – again the only good explanation we can think of – is to assume that causal relations are *directed*: In what follows "X→Y" expresses that X exerts a causal influence on Y "directly", i.e. unmediated relative to the given set of variables V. The way this direct causal influence is physically realized is left unspecified in TCN. However, two assumptions are required that are precisely formulated in sec. 2.3 and *informally* stated as follows:

---

[6] In the sun-tower-shadow example we can infer every Y-value from every X-value for every Z-value by the equation Y = Z/tan(X). In other examples, X and Y become only correlated when conditionalizing on *certain* values of the common effect Z.

Productivity (P): "Ceteris absentibus" (i.e. in the absence of intervening causal influences)[7] X→Y implies a probabilistic dependence between X and Y, and

Markov-causality (C): Probabilistic dependencies are the result of directed causal connections, which transmit probabilistic influence from causes to effects, but not from effects to causes.

We can now explain screening off *and* linking up phenomena as follows. In both cases, Z mediates between X and Y. So we have three possible directed causal structures as candidates for explaining these phenomena:

(a) X→Z→Y (or X←Z←Y) ("chain"): Z is an intermediate cause (between X and Y).

(b) X←Z→Y ("fork"): Z is a common cause (of X and Y).

(c) X→Z←Y ("collider"): Z is a common effect (of X and Y).

The first two arrangements explain screening off; the third one explains linking up.

Explaining screening off:

(a) Chain (X→Z→Y): Y depends on X because a change of X-values causes a change of Z-values which, in turn, causes a change of Y-values (DEP(X,Y)).

(b) Fork (X←Z→Y): Y depends on X because changes of X-values are caused by changes of Z-values which also cause changes of Y-values (DEP(X,Y)).

In case (a) as well as case (b), X-value variations can lead to Y-value variations only due to Z-value variations; thus fixing Z's value renders X and Y independent (INDEP(X,Y|Z)).

The logical structure of both explanations is as follows: From X→Z in case (a), or from Z→X in case (b), we infer DEP(X,Z) by (P); likewise we infer DEP(Z,Y) from Z→Y and (P).

---

[7] More precisely, we must deactivate all other causal connections between X and Y and other causal influnces on Y; see sec. 2.3, (9).

INDEP(X,Y|Z) follows directly from (C) and causal structures (a) and (b). For explaining DEP(X,Y) we must additionally assume that the dependencies DEP(X,Z) and DEP(Z,Y) are *transitive* in the sense that they result in DEP(X,Y). To this end it is necessary and sufficient that the causal models (a) or (b), resp., satisfy the following condition of dependence transitivity (DT):[8] $\exists x,y: \Sigma_{z\in Val(Z)}P(y|z)\cdot P(z|x) \neq \Sigma_{z\in Val(Z)}P(y|z)\cdot P(z)$.

DEP(Y,Z) and DEP(Z,X) imply that $P(y|z) \neq P(y|\neg z)$ holds for some y and z, and that $P(z|x) \neq P(z|\neg x)$ holds for some x,z. So condition (DT) can only be violated when positive and negative changes of certain terms "P(y|z)·P(z|x)" in the left sum, compared to the corresponding terms "P(y|z)·P(z)" in the right sum, cancel out to zero. Hence a violation of condition (DT) occurs only in rare cases that correspond to non-robust (unfaithful) causal scenarios. In other words, condition (DT) is not only needed to explain screening off, but also intrinsically plausible.

Explaining linking up:

(c) Collider (X→Z←Y): Y doesn't depend on X because a change of X-values causes a change of Z-values which, however, isn't accompanied by a change of Y-values because value-changes are not transmitted from an effect to its cause. Fixing Z to certain values will render X and Y dependent (DEP(X,Y|Z)), as explained in the sun-tower-shadow example (5).

The logical structure of this explanation starts again with the observation that by (P), X→Z and Z←Y imply DEP(X,Z) and DEP(Z,Y), respectively. By (C), no probabilistic influence of a cause X on its effect Z can be transmitted to Z's other cause Y; so "ceteris absentibus" X and Y are probabilistically independent, i.e. INDEP(X,Y). In order to explain DEP(X,Y|Z) it suffices to

---

[8] *Proof:* By probability theory we have (a) $P(y|x) = \Sigma_z P(y|x,z)\cdot P(z|x)$ and (b) $P(y) = \Sigma_z P(y|z)\cdot P(z)$. The sum in (a) equals (c) $\Sigma_z P(y|z)\cdot P(z|x)$ by condition (C) of sec. 2.3, since Y is not d-connected with X given Z. It follows that $P(y|x) \neq P(y)$ holds exactly if the two sums in (c) and (b) are unequal. Q.E.D.

assume that the causal model (c) satisfies the following condition of dependence overlap (DO):[9] $\exists x,y,z$: $Dep(y,z) \wedge Dep(z,x)$.[10]

A violation of condition (DO) would not be robust against minimal changes of the involved conditional probabilities. So, again, the satisfaction of (DO) is not only needed to explain linking up, but also intrinsically plausible.

With help of directed arrows, we are able to explain even certain *non-robust* cases of screening off and linking up. A non-robust scenario in which Z screens off X from Y could, for example, be explained by the causal structure in fig. 3(a): the positive conditional dependence due to $X \rightarrow Z \leftarrow Y$ and the negative dependence due to $X \rightarrow Y$ cancel out to zero. Thus, DEP(X,Y) and INDEP(X,Y|Z), though Z is a common effect of X and Y. In sec. 3.2 we call this situation *un-*

---

[9] *Proof:* By probability theory, (a) $P(x|y) = P(x|y,z)\cdot P(z|y)+P(x|y,\neg z)\cdot P(\neg z|y)$ and (b) $P(x) = P(x|z)\cdot P(z)+P(x|\neg z)\cdot P(\neg z)$. By INDEP(X,Y) we have $P(y|x) = P(y)$. So the sums in (a) and (b) must be equal. These sums are weighted averages, with the weights in the sum in (a) being $P(z|y)$ and $P(\neg z|y) = 1-P(z|y)$, and the weights in the sum in (b) being $P(z)$ and $P(\neg z) = 1-P(z)$. By (DO) we have (i) $P(z|y) \neq P(z|\neg y)$ and (ii) $P(x|z) \neq P(x|\neg z)$. It follows from (i), (ii), and the laws of weighted averages that the two sums in (a) and (b) would have to be different if INDEP(x,y|Z), i.e. $P(x|y,z) = P(x|z)$ and $P(x|y,\neg z) = P(x|\neg z)$, would hold. For if $a \neq b$ and $w \neq w'$, then $a\cdot w+b\cdot(1-w) = (a-b)\cdot w+b \neq a\cdot w'+b\cdot(1-w') = (a-b)\cdot w'+b$. Thus either $P(x|y,z) \neq P(x|z)$ or $P(x|y,\neg z) \neq P(x|\neg z)$ must hold, which gives us DEP(X,Y|Z). Q.E.D.

[10] Condition (DO) is sufficient but not necessary for DEP(Y,X|Z) in linking up cases (it can be shown that a necessary condition for DEP(Y,X|Z) is $\exists x,y,z$: DEP(y,z|x)∧DEP(z,x|y)). On the other hand, condition (DO) is not sufficient but necessary for DEP(Y,X) in screening off cases (we are indebted to an anonymous reviewer for pointing this out to us). Here is a *proof* of the necessity-claim by contraposition ($\neg$(DO) $\Rightarrow$ INDEP(X,Y)). Assume that (DO) fails (in one of the causal structures $X \rightarrow Z \rightarrow Y$, $X \leftarrow Z \leftarrow Y$, or $X \leftarrow Z \rightarrow Y$). Thus $P(z|y) = P(z)$ holds for every z with $P(x|z) \neq P(x|\neg z)$. Let Zy be the set of Z's values that are Y-independent but X-dependent, and Zx the set of Z's values that are X-independent but Y-dependent. Then $P(x|y) = \Sigma_z P(x|z)\cdot P(z|y) = \Sigma_{z \in Zy} P(x|z)\cdot P(z|y) + \Sigma_{z \in Zx} P(x|z)\cdot P(z|y)$ = (by our assumptions) $\Sigma_{z \in Zy} P(x|z)\cdot P(z) + \Sigma_{z \in Zx} P(x)\cdot P(z|y) = \Sigma_{z \in Zy} P(x,z) + P(x)\cdot\Sigma_{z \in Zx} P(z|y)$ = (*) $P(x,z \in Zy) + P(x)\cdot P(z \in Zx|y)$. By our assumption, $P(y|z \in Zy) = P(y)$ holds, which implies $P(z \in Zy|y) = P(z \in Zy)$. This implies $P(z \in Zx|y) = P(z \in Zx)$ (via $P(z \in Zx|y) = 1-P(z \in Zy|y) = 1-P(z \in Zy) = P(z \in Zx)$), which in turn implies $P(x)\cdot P(z \in Zx|y) = P(x)\cdot P(z \in Zx) = P(x, z \in Zx)$. So we continue as follows $P(x|y) = \ldots$ (*) $= P(x,z \in Zy) + P(x,z \in Zx) = P(x)$. Thus, INDEP(x,y).

faithfulness due to canceling paths. Unfaithful independencies are not robust, since small changes of the involved conditional probabilities turn them into dependencies. An analogous alternative explanation can be given for the non-robust linking up case in fig. 3(b), in which the positive dependence due to X→·Z→·Y and the negative dependence due to X→·Y cancel out to zero.

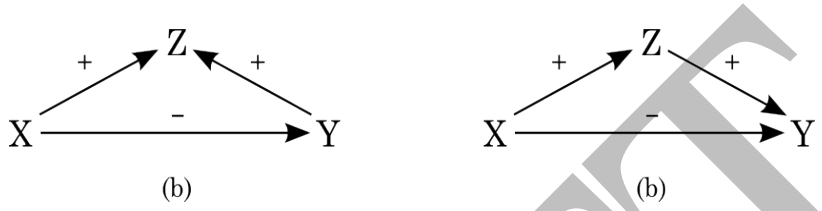


Fig. 3: *Unfaithfulness due to canceling paths*: (a) explains non-robust screening off (DEP(X,Y) and INDEP(X,Y|Z)); (b) explains non-robust linking up (INDEP(X,Y) and DEP(X,Y|Z)).

We finally remark that two independent variables X and Y may not only be linked up by common effects, but also by effects of common effects. An example is illustrated in fig. 4: assuming iterated dependence transitivity $\Sigma_z P(y|x,z',z)\cdot P(z|x,z') \neq \Sigma_z P(y|z',z)\cdot P(z|z')$, X and Y will be dependent conditional on Z':

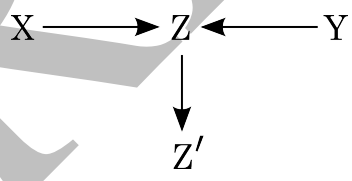


Fig. 4 X and Y are linked up by Z'.

Summarizing, we argued that the best available explanation for robust as well as for non-robust cases of screening off and linking up is to assume that probabilistic dependence between variables is mediated by directed binary causal relations which obey (C) and (P). Thus, causation as characterized by (C) and (P) can be justified as ontologically real by an inference to the best available explanation (IBE).

We see the major advantage of the proposed justification in the fact that it neither presuppos-

es advanced concepts of physics nor strong metaphysical assumptions. It rather justifies causality on the basis of ordinary phenomena in everyday life. In particular, we experience screening off and linking up in all kinds of *purposeful* actions. Here our actions (A) realize certain means (M) in order to produce certain purposes (P) (A→M→P); so DEP(A,P), but M screens off A from P, i.e. our actions cannot reach their purposes without realizing certain means. Moreover, if a purpose P can be achieved by two independent means M_1 and M_2, then the achievement of purpose P links up M_1 and M_2 (if M_1 was not applied, M_2 has been applied). These facts help to explain why causality is an inborn reasoning mechanism of homo sapiens which is closely connected to interventions.

Let us finally compare our IBE justification strategy with the well-known *fork asymmetry* argument. This argument, which goes back to Reichenbach (1956, 159-61) and has been elaborated by Papineau (1992), runs as follows: Assume X,Y are two events both correlated with a third event Z. Then either (a) X and Y are mutually correlated and Z screens them off: then Z is a common cause of X and Y. Or (b) X and Y are uncorrelated: then Z is a common effect of X and Y. Note that this justification strategy has gaps. As Papineau (1992, 240) observes, the argument doesn't work if X and Y can causally reach each other; Reichenbach excluded this case by assuming that X and Y are temporally simultaneous. Another gap is the third possible case (c) in which X and Y are correlated but Z doesn't screen them off, because X, Y, and Z are joint effects of a common cause C. In contrast, our proposed justification strategy doesn't suffer from these restrictions. It is not based on the fork asymmetry, but on the asymmetry between screening off and linking up, which is not considered by Reichenbach or Papineau.

2.3 The core axioms of TCN: causal Markov (d-connection) and productivity (minimality)

We now present the concise statement of the axioms of d-connection (C) and productivity (P) that have been justified by an IBE in sec. 2.2. (C) and (P) constitute TCN's core and are tradi-

tionally expressed by the equivalent causal Markov condition (M) and the minimality condition (Min), respectively (SGS 2000, sec. 3.4.1-2). We prefer (C) and (P) over (M) and (Min) because they are philosophically more transparent and are better suited for investigating TCN's empirical content than (M) and (Min). Before we explicitly state (C) and (P), we have to introduce the following notions. A *causal graph* (or *structure*) is a pair $(\mathcal{V},\mathcal{E})$, where $\mathcal{V}$ is a set of variables (the "vertices"), and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is a set of directed arrows $X_i \rightarrow X_j$ (i.e. ordered pairs, the "edges"). A graph $(\mathcal{V},\mathcal{E})$ together with a probability distribution P over $\mathcal{V}$ is called a *causal model* (or *system*) $(\mathcal{V},\mathcal{E},P)$. Causal structures and systems are parts of the world, while causal graphs (CGs) and models (CMs) are conceptual representations of causal structures and systems, respectively. Some further important notation:

- $X \rightarrow Y$: X is a *direct cause* of Y (Y is a *direct effect* of X)
- $X \rightarrow\rightarrow Y$: X is a (direct or indirect) *cause* of Y (Y is an *effect* of X), i.e. there is a directed path $X \rightarrow Z_1 \rightarrow \ldots \rightarrow Z_n \rightarrow Y$ from X to Y.
- $X - Y$: $X \rightarrow Y$ or $X \leftarrow Y$, i.e. X and Y are *adjacent*.
- $X_1 - - X_n$: a *path* $X_1 - \ldots - X_n$ between X_1 and X_n; this path *connects* X_1 and X_n and the variables X_i $(1 \leq i \leq n)$ *lie* on this path.

– If X_i lies on a path π, then X_i is called (i) a common cause, (ii) an intermediate cause, or (iii) a common effect on π iff (i) ←X→, (ii) →X→ or ←X←, or (iii) →X←, respectively, is part of π.

The principle of d-connection says that every (conditional) probabilistic dependence between two variables X and Y is the result of some causal path connecting them. The correct formulation of this principle has to account for all possible combinations of screening off and linking up along *all* paths connecting X and Y. If path π connects X and Y in a graph ($\mathcal{V}$,$\mathcal{E}$), then π can generate probabilistic dependence conditional on a (possibly empty) subset of variables **U** ⊆ $\mathcal{V}$–{X,Y} only if *no* common or intermediate cause on π is in **U** and *all* common effects on π are in **U** or have an effect in **U**. If X and Y are connected in a graph ($\mathcal{V}$,$\mathcal{E}$) by several paths, then X and Y become dependent *iff at least one* of these paths generates a probabilistic dependence. These considerations are summarized in the following axiom of d-connection (C):

(6) *Axiom of d-connection (C):* Every physically possible CM ($\mathcal{V}$,$\mathcal{E}$,P) [in an intended domain] satisfies the *condition of d-connection (C)*, which is defined as follows:

For all X,Y∈V and **U** ⊆ $\mathcal{V}$–{X,Y}: *If* DEP(X,Y|**U**), *then* X and Y are *d-connected* given **U**

in the following sense:

X and Y are connected by some path π such that no intermediate or common cause on π is in **U**, while every common effect on π is in **U** or has an effect in **U**. (In this case we say that X and Y are d-connected given **U** *by* π).

Note that in (6) we distinguish between the *definition* of the d-connection *condition* (which is a property that a causal model may or may not have) and the corresponding *axiom* of d-connection, which states that this condition holds for all physically possible causal models in an intended domain. In what follows, when we simply write "(C)" we always mean the *condition* of d-connection, while when referring to the axiom of d-connection we will explicitly write "axiom (C)". We assert axiom (C) for physically possible (rather than only for actual) causal models since the causal systems in our world and their probability distributions may change over time.

Condition (C) entails the following well-known principle of

(7) *Unconditional dependence:* If DEP(X,Y), then X and Y are connected by a directed or common cause path (i.e. d-connected given ∅).

If X and Y are connected by a path π, then **U** is said to *activate* π iff X and Y are d-connected given **U** by π, while **U** *blocks* π iff X and Y are not d-connected given **U** by π. If X and Y are not d-connected given **U**, X and Y are said to be *d-separated* by **U**. The concepts of d-separation and d-connection have been developed by Pearl (1988, 117). (C) asserts an *implication* from probabilistic dependence to d-connection – or, in contraposed form, an implication from d-separation

to independence, which is the formulation used by Pearl (1988, 119; 2000, th. 1.2.4). (C) is equivalent with two famous conditions, the causal Markov condition and Markov-compatibility (cf. Pearl 2009, 16; SGS 2000, 29f). In the following par(X) is the set of X's parents, i.e. the direct causes of X in the given causal graph:

(8) *Definition of the causal Markov condition (M) and of Markov-compatibility (MC)*:

(8.1) ($\mathcal{V}$,$\mathcal{E}$,P) satisfies (M) *iff* every X$\in\mathcal{V}$ is independent of all of its non-effects (different from X) conditional on its parents, i.e. INDEP(X,**U**|par(X)) holds for every subset **U** $\subseteq$ $\mathcal{V}$–{X} of non-effects of X.

(8.2) ({X_1,…,X_n},$\mathcal{E}$,P) satisfies (MC) *iff* $P(X_1,\ldots,X_n) = \Pi_{1\leq i\leq n}P(X_i|par(X_i))$.

SGS (2000, sec. 3.4.1-2) prefer (M) as the core of TCN. The equivalence of (M) and (MC) is well known (e.g. Pearl 2009, 19, th. 1.2.7); the product at the right of "=" in (8.2) is called the *Markov-factorization* of the CG ({X_1,…,X_n},$\mathcal{E}$,P). Deeper is the equivalence between (M) and (C), which has been demonstrated by Verma (1986), Pearl (1988, 119f, th. 9, cor. 4), and Lauritzen et al. (1990, 50), who call (C) the *global* and (M) the *local* Markov condition (see also SGS 46, th. 3.3; Pearl 2009, 18, th. 1.2.4). For an intuitive grasp of the relation between (C) and

(M), observe that (in its contraposed form) (C) asserts a (conditional) independence for *all* d-separation relations of a causal graph, while (M) asserts such an independence only for the d-separation relations between a variable X and its non-effects conditional on its parents; the other independencies follow from these as probabilistic consequences. On this reason, (C) expresses the full content of the causal Markov condition in a much more direct way than (M), whence we prefer reference to (C) over reference to (M) in the first core axiom of TCN.

Note that the equivalence of (C) and (M) holds only for *acyclic* causal graphs, i.e. CGs not containing cyclic paths $X \rightarrow\rightarrow X$.[11] Moreover, the equivalence of (C) with (MC) presupposes finiteness of $\mathcal{V}$:

Theorem 1: For every acyclic finite CM: (C), (M), and (MC) are pairwise equivalent.

Under the assumption that causation is forward-directed in time, cyclic CGs are impossible. However, directed causal dependencies appear to hold also between temporally coexisting properties of stationary systems: the length of a pendulum, for example, is a cause of the pendulum's frequency. To avoid misunderstandings, a causal arrow between two coexisting variables of a stationary systems should not be understood as an instantaneous physical interaction. It rather refers to a causal processes that goes on in a given finite time *interval*. More importantly, it implies an intervention asymmetry: one can change the value of the effect variable by changing the value of the cause variable, but not vice versa.

In special cases, the causal dependencies between coexisting properties of stationary systems may be cyclic. Examples are *self-regulatory systems* containing feedback loops, e.g.: outside

[11] An example of a cyclic CG violating (C) $\Leftrightarrow$ (M) is found in Spirtes et al. (1993, 359f).

temperature→room temperature $\rightleftarrows$ thermostat. (On the explained reason such causal cycles are not in conflict with the temporal forward-directedness of causal processes.) While (M) makes only good sense for acyclic graphs, (C) is also reasonable for cyclic graphs (cf. Spirtes et al. 1993, 359). In this paper, however, we focus on acyclic CGs.

The distributions P(X|par(X)) are called the causal model's *parameters*.[12] In acyclic models, these parameters can be varied independently from each other without destroying the independencies entailed by (C).[13]

We understand axiom (C) and the equivalent axiom (M) as synthetic (i.e. not analytically true) principles whose content can be true or false in the realistic sense.[14] These axioms assert that (C) holds for most physically possible worlds or physically closed systems (universes). Moreover, (C) provably holds for every *subsystem* of a (C)-satisfying causal system ($\mathcal{V}$,$\mathcal{E}$,P) that is *causally sufficient*, i.e. that doesn't omit any true and non-degenerate common cause of variables in $\mathcal{V}$ (cf. SGS, 22).

SGS (2000, 29) speak of conditions (C) and (M) likewise as of "axioms", but present them in the form of definitions; Pearl (2009) only states the definitions. The explicit formulation of axioms impels us to critically reflect the *problem of generality*: do really *all* correlations result from

---

[12] These parameters can equivalently be formulated as functions $X = f(par(X))+U_X$ together with a random distribution P over mutually independent error variables U_X (Pearl 2000, 44). If the causal influences are non-interactive and linear, one can factorize the parents' influences in the form of a structural equation model, $X = \Sigma_{P_i \in par(X)} c_i \cdot P_i + U_X$ (SGS 14f).

[13] This follows from (MC).

[14] Also Pearl (2000, preface) and SGS (ch. 3.4-5) support a realistic understanding of causal relations.

causal connections? The justification of causality by an IBE in sec. 2.2 works only for correlations that participate in relations of screening off or linking up. We argued that all correlations that can be utilized by means of interventions participate in screening off and/or linking up relations – however, not all correlated variables can be intervened on. Possible failures of the causal Markov condition have been discussed in the context of EPR-correlations in quantum mechanics, in which the correlated states of two entangled particles are not screened off by common causes, though they cannot be explained by a direct causal connection due to relativity theory (cf. Hausman 1998, 252; Healey 2009). Cartwright (2007, 122) argues that similar problems may even arise in ordinary (macroscopic) domains. Well-taken defenses against these objections have been given by the proponents of TCN (SGS 2000, 59-63; Pearl 2009, 62, Hitchcock 2010), though not all problems are solved by these defenses and the debate is ongoing. In this paper we don't take up a stance on whether (C) is strictly general or holds only in certain domains. To account for the latter possibility we added "[in an intended domain]" in axiom (C). Yet we will have something significant to say about the arguments concerning TCN's confirmation status in sec. 3.1, when we discuss the philosophically puzzling consequences of the fact that TCN's core axioms are empirically empty.

Axiom (C) asserts that a probabilistic dependence implies a causal connection. The second core axiom, (P), asserts that the other direction, from causal connection to probabilistic dependence (or from independence to causal separation), holds under *special* conditions: a *direct* causal connection implies *ceteris absentibus* a probabilistic dependence. (P) holds only under certain conditions because unfaithful structures due to *canceling paths* (fig. 3, sec. 2.2) cannot be excluded a priori. Another kind of unfaithfulness is unfaithfulness due to *canceling parents*. It arises when two causal parents interact in such a way that their influences on their joint child cancel each other (see fig. 5). Pearl (1988, 256) gives an illustrative example: a bell Y rings iff two randomly tossed coins X and Z both land heads or tails: $Y = 1 \Leftrightarrow (X = Z)$. Thus, $X = 1$ influences Y

= 1 positively if Z = 1 and negatively if Z = 0, where these two influences cancel each other, since P(Y = 1|X = 1) = P(Y = 1|Z = 1) = P(Y = 1) = 0.5.

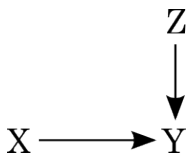


Fig. 5: *Unfaithfulness due to canceling parents*: INDEP(X,Y), INDEP(Z,Y), INDEP(X,Z), but DEP(X,Y|Z) and DEP(Z,Y|X)

We can isolate X's causal influence on Y from the influence of possibly canceling paths or parents by blocking this influence via conditionalization. In acyclic causal structures this can be done by conditionalizing on Y's parents (different from X). So we explicate the axiom of productivity as follows, where (as in the case of (C)) we distinguish between the condition (P) and the axiom (P):

(9) *Axiom of productivity (P):* Every physically possible CM ($\mathcal{V}$,$\mathcal{E}$,P) [in an intended domain]

satisfies the *condition of productivity (P)*, which is defined as follows:

DEP(X,Y|par(Y)–{X}) holds for all X→Y in $\mathcal{E}$.

Lemma 1: For every acyclic CG ($\mathcal{V}$,$\mathcal{E}$,P) satisfying (C): If X→Y in $\mathcal{E}$, then:

(1.1) par(Y)−{X} d-separates X from Y in ($\mathcal{V},\mathcal{E}$−{X→Y}).

(1.2) If DEP(X,Y|**U**) holds for some **U** ⊇ par(Y)−{X} that d-separates X from Y in ($\mathcal{V},\mathcal{E}$−{X→Y}), then DEP(X,Y|par(Y)−{X}).

Lemma 1.1 justifies definition (9) for acyclic graphs (proof see appendix). (In cyclic graphs, such as X→Y→Z, the conditioning set may have to include further nodes, e.g. X's parent Z, in order to isolate the dependence generated by X→Y.) Lemma (1.2) provides additional information about condition (P) (note that the only-if direction of (1.2) holds trivially).

In distinction to the first core axiom, axiom (P) is justified by a methodological requirement: in order to be empirically significant, causal arrows must be responsible for at least some (conditional) probabilistic dependencies; causal arrows without empirical effects are eliminated by Ockham's razor.

Given (C), (P) is equivalent with the well-known minimality condition (Min) (SGS 31):

(10) *Definition:* A (C)-satisfying CM ($\mathcal{V},\mathcal{E}$,P) is minimal (i.e. satisfies (Min)) *iff* no arrow can be omitted from $\mathcal{E}$ without violating condition (C), i.e. every submodel ($\mathcal{V},\mathcal{E}'$,P) with $\mathcal{E}'$

$\subset \mathcal{E}$ violates (C).

Theorem 2: For all finite acyclic CMs satisfying (C): (Min) and (P) are equivalent.

The notion of productivity and theorem 2 are new; its proof is stated in the appendix (the condition of finiteness in theorem 2 could be relaxed, but only on the cost of a much more complicated proof). The advantage of (P) over (Min) is twofold. First, (Min) tells us only that every arrow X→Y is needed to explain some dependence within the given CM, while (P) states this dependence explicitly. Second, (P) is independent of (C), while (Min) presupposes (C). This does not only hold for minimality as defined in (10), but also for Zhang and Spirtes' (2011, 182) definition: "If (C) holds in the given CM, then (C) holds in no $\mathcal{E}$-contraction of this CM." According to this definition every model violating (C) would trivially be minimal, which is certainly not intended. In contrast, (P) does also make sense in case of (C)-violating CMs, such as quantum-mechanical models with common causes not screening off their effects. While (Min) cannot be sensibly applied to such cases, (P) may either hold or be violated.

3. Empirical content of TCN

In sec. 2 we showed that TCN is needed to explain screening off and linking up. In order to be empirically significant, these explanations should not be entirely post facto, i.e. TCN should not be compatible with every logically possible probability distribution. In other words, TCN should have empirical (or non-theoretical) content.

In investigating TCN's empirical content we follow the analogy between causality in TCN and force in classical physics mentioned in sec. 1. As the total force law (sum of forces = mass · acceleration) and the actio-equals-reactio law constitute the core of classical physics, axioms (C) and (P) constitute the core of TCN. But there are further general principles, such as faithfulness (F), the noise condition (EN), and temporal forward-directedness (T), which are introduced in sec. 3.1 and sec. 3.2. These principles constitute *extended versions* of TCN just like the law of gravitational or frictional force constitutes extended versions of Newtonian physics.

In regard to the question of empirical content it is important to distinguish between the empirical content of the *general* theory TCN and the empirical content of *particular* CMs of TCN. CMs describe particular *applications* of TCN in TCN's theoretical language, i.e. in terms of causal models. This is analogous to the distinction between the general force theory of classical physics and particular force models such as a sun-planet-system. Axiom (C) generates empirical content for particular CMs by entailing probabilistic independencies. The inverse inference from empirical probability distributions ($\mathcal{V}$,P) to CMs is more difficult, since it faces the problem of empirical underdetermination: the same probability distribution may be explainable by more than one causal structure satisfying TCN's core axioms. A lot of work in the theory of causal nets has been concerned with this *causal inference* problem. But even under conditions under which this inference problem is solved and the inference from a given distribution to a CM is unique, TCN's causal explanations could still be without empirical content.

In other words, the problems of causal inference and empirical content are largely independent. According to our knowledge, the problem of TCN's empirical content has so far not been investigated by logical means. TCN has empirical content *iff* it *excludes* some analytically possi-

ble empirical models. A *non-theoretical* model is a pair ($\mathcal{V}$,P) of a set of variables $\mathcal{V}$ together with a probability distribution P over $\mathcal{V}$. A non-theoretical model ($\mathcal{V}$,P) is *empirical* if $\mathcal{V}$ is a set of empirically measurable variables. An empirical (or non-theoretical) model ($\mathcal{V}$,P) is called an empirical (or non-theoretical) *submodel* of a CM ($\mathcal{V}'$,$\mathcal{E}'$,P') iff $\mathcal{V} \subseteq \mathcal{V}'$ and P = P'$\upharpoonright \mathcal{V}$ (the restriction of P' to $\mathcal{V}$).[15] We also say that ($\mathcal{V}'$,$\mathcal{E}'$,P') *expands* ($\mathcal{V}$,P). We define:

(11) *Empirical (non-theoretical) content for TCN:* A version of TCN has empirical (or non-theoretical) content *iff* there exists a logically possible empirical (or non-theoretical) model that cannot be expanded to a CM satisfying this version of TCN.

If TCN did *not* have empirical content, the content of all particular CMs would be entirely ex-post. For any empirical model ($\mathcal{V}$,P) one could then invent a "causal explanation" in accordance

[15] Empirical submodels correspond to what is called "partial (potential) models" in structuralist philosophy of science (cf. Balzer et al. 1987; Sneed 1971, ch. 3).

with TCN, i.e., a TCN-model ($\mathcal{V}$,$\mathcal{E}$',P') that expands ($\mathcal{V}$,P). If this were the case, TCN would be exposed to the objection of being superfluous "causal metaphysics". It would then be impossible to predict a new probabilistic (in)dependence R_{n+1} which is not probabilistically entailed by the already known (in)dependencies $R_1,\ldots,R_n$ by means of TCN.

3.1 Empirical content of TCN's core: causal Markov, productivity, and acyclicity

We start with the question whether TCN's core has empirical content. Our result, stated in theorem 3, is negative: (C)+(P) alone don't have empirical content, not even when adding acyclicity:

Theorem 3: Every analytically possible empirical (or non-theoretical) model ($\mathcal{V}$,P) can be expanded to an acyclic CM ($\mathcal{V}$,$\mathcal{E}$',P') satisfying (C) and (P).

Theorem 3 follows from a well-known property of Bayesian nets (cf. Pearl 2009, 14). For every ordering $X_1,\ldots,X_n$ of the variables in $\mathcal{V}$, (*) $P(X_1,\ldots,X_n) = \Pi_{1\leq i\leq n}P(X_i|X_1,\ldots,X_{i-1})$ is probabilistically valid, where (*) holds for all value-instantiations $x_1,\ldots,x_n$ of $X_1,\ldots,X_n$. By excluding those X_j from which X_i ($1 \leq j \leq i-1$) is probabilistically independent in each term $P(X_i|X_1,\ldots,X_{i-1})$ in (*), one obtains the so-called "Markovian parents" $mp(X_i)$ of X_i w.r.t. the ordering $X_1,\ldots,X_n$. (Note

that the Markovian parents are defined for *every* given ordering of variables, so they need not be the "true" causal parents.) By drawing a CG whose arrows point from each member of a non-empty parent-set $mp(X_i)$ to X_i, one obtains an acyclic minimal (C)-satisfying CM.

Technically, theorem 3 is unspectacular. Its philosophical consequences, however, deserve a critical reflection, in particular w.r.t. the philosophical debate on the empirical confirmation status of the equivalent causal Markov condition (M). Proponents of TCN have argued that (M) is satisfied by all (or most of all) known empirical and/or technical systems (cf. SGS 2000, 29; Pearl 2009, 62f; Hitchcock 2010, sec. 3.3). However, since TCN's core axioms are empirically empty (theorem 3), it is impossible to confirm axiom (C) without additional assumptions – but no additional assumptions are stated in the cited passages. The same problem applies to critics of axiom (C): to turn their examples into counterexamples to axiom (C), they must make further assumptions about causality. In the example of a common cause structure $X \leftarrow Z \rightarrow Y$ in which Z doesn't screen off X from Y (e.g. Cartwright 2007, 122), these additional assumptions are usually causal sufficiency (no hidden common causes of X and Y) and separation (the joint effects are not causes of each other). Even if such additional assumptions were to be explicitly stated, their content when added to TCN is not obvious. We therefore think that investigating TCN's content is an important task for the TCN research program.

Is it a problem that TCN's core is empirically empty? Not necessarily. It is rather typical for scientific theories that their cores are empty. Sneed (1971, 126) has demonstrated with scrutiny that, for example, the core of classical physics, the total force law, is empirically empty. For every system of (point) masses with given accelerations one can construct force functions that satisfy the total force law. However, it is also well-known that the empirical content of general classical physics abruptly increases when special force laws (e.g. the law of gravitational force) are added. Do we meet a similar situation in case of TCN? This is the question of the next two sec.

3.2 Empirical content of faithfulness and noise assumption

(C) asserts that probabilistic dependence implies d-connection. (P) asserts the inverse implication relation under very restricted conditions. The *full* content of the inverse implication is called the *faithfulness condition* (cf. SGS 2000, 31):

(12) *Definition of the faithfulness condition (F):* ($\mathcal{V}$,$\mathcal{E}$,P) satisfies (F) *iff* ($\mathcal{V}$,$\mathcal{E}$,P) satisfies the converse of (C): *if* X and Y are d-connected given $\mathbf{U} \subseteq \mathcal{V}$–{X,Y}, *then* DEP(X,Y|**U**).

In other words, a CM is faithful iff P verifies *only* those probabilistic independence relations that are implied by (C). SGS (ibid.) define "faithfulness" as the conjunction of (C) and (F). We prefer our definition (which Zhang and Spirtes 2008, 247 also use), because it logically separates (F) from (C). This is important because there are several possibilities for (F) to be violated. For this reason we don't introduce (F) as an axiom, but a probabilistic weakening of (F) (see below). It is easily seen that:

(13) Faithfulness implies productivity.

For assume ($\mathcal{V}$,$\mathcal{E}$,P) does not satisfy (P). Then there is an X→Y in $\mathcal{E}$ such that IN-DEP(X,Y|par(Y)–{X}). But since X and Y are d-connected given par(Y)–{X} in ($\mathcal{V}$,$\mathcal{E}$),

($\mathcal{V},\mathcal{E}$,P) is not faithful.

Faithfulness is much stronger than productivity. Contrary to (P), (F) has various exceptions. (F) may be violated because of special and usually rare features of given probability distributions. We call an independence INDEP(X,Y|**U**) in a CM ($\mathcal{V},\mathcal{E}$,P) an *unfaithful independence* iff **U** d-connects X and Y in ($\mathcal{V},\mathcal{E}$). For formulating an empirically tenable "axiom" of faithfulness, it is useful to distinguish three types of unfaithful independencies:

1.) Cancelation unfaithfulness: This type has been explained in sec. 2.2. It has two subtypes: (1.1) Unfaithfulness due to *canceling paths* (fig. 3(a) and 3(b)), and (1.2) unfaithfulness due to *canceling parents* (fig. 5).

2.) Determinism unfaithfulness: The value x of a variable X depends deterministically on a set of values **w**, abbreviated as "det(x:**w**)", iff P(x|**w**)∈{0,1} holds; x depends deterministically on a set of variables **W** iff det(x:**w**) holds for all **W**-values **w**. Determinism unfaithfulness can arise when X and Y are d-connected by a path π given **U**, but INDEP(X,Y|**U**) holds because π carries a variable Z* (possibly identical with X or Y) and all values z* of Z* which transmit probabilistic influence from X to Y are deterministically dependent on some variable Z∈**U** (SGS 2000, 53ff). An example is the causal model X→Z→Y←Z' in which det(y:Z') holds for all y with DEP(X,y); in this case we get INDEP(X,Y|**U** = {Z'}), though X and Y are d-connected given **U**.

3.) Intransitivity unfaithfulness: Here the unfaithfulness independence is produced by de-

pendencies between adjacent pairs of variables that are not transitive. In the simplest case they arise from a non-overlap of dependence-sensitive values: here all adjacent nodes on a path X_1–…–X_n are dependent (DEP(X_i,X_{i+1}|**U**)), but this dependence is not transmitted from X_1 to X_n (INDEP(X_1,X_n|**U**)) because the value-pairs (x_i,x_{i+1}) in which the adjacent variables are dependent don't yield an overlapping chain (x_1,…,x_n). A simple example is a chain X→Z→Y or fork X←Z→Y in which Z has four values, but X depends on Z only over $\{z_1,z_2\}$ and on Y only over $\{z_3,z_4\}$, i.e. P(z_i|X) ≠ P(z_i) *iff* i∈{1,2} and P(z_i|Y) ≠ P(z_i) *iff* i∈{3,4}). In this case, ¬∃x,y,z: DEP(x,z)∧DEP(z,y) holds, i.e. condition (DO) of sec. 2.2 is violated, which implies INDEP(X,Y), as proved in fn. 10. Intransitive dependencies may also arise in a common effect structure X→Z←Y: here ¬∃x,z,y: DEP(x,z)∧DEP(z,y) implies INDEP(X,Y|Z) (see the proof in fn. 10). More complicated cases of intransitive dependencies may obtain in chain or fork structures which arise in spite of condition (DO) (Naeger (this volume) calls them "internal canceling paths"). Moreover, intransitive dependencies may also arise for chains of arbitrary length (Zhang and Spirtes 2008, 253).

It is easy to see that axiom (C) together with the faithfulness condition (F) have empirical content. A result of this kind can be found in Zhang and Spirtes (2008, 253), though not in terms of content, but in terms of "detectable kinds of unfaithfulness". The authors define the unfaithfulness of a given (C)-satisfying causal model ($\mathcal{V}$,$\mathcal{E}$,P) as *detectable* iff its non-theoretical submodel ($\mathcal{V}$,P) cannot be expanded into a causal model ($\mathcal{V}$,$\mathcal{E}'$,P) satisfying (C)+(F). Zhang and Spirtes' theorems imply the following results for the empirical content of (C)+(F):

Theorem 4: (C)+(F) have empirical (or non-theoretical) content: No empirical (or non-theoretical) model ($\mathcal{V}$,P) with {X,Y,Z} ⊆ $\mathcal{V}$ verifying the logically possible (in)dependence relations in (4.1) or (4.2) can be expanded to a CM ($\mathcal{V}$,$\mathcal{E}$',P'):

(4.1) (a) ∀**U** ⊆ $\mathcal{V}$–{X,Y}: DEP(X,Y|**U**)∧DEP(Y,Z|**U**), and (b) there exist two distinct sets **W**,**W**' ⊆ $\mathcal{V}$–{X,Z} with Y∈**W** but Y∉**W**' which screen off X from Z.

(4.2) INDEP(X,Y), INDEP(Y,Z), INDEP(X,Z), DEP(X,Y|Z), DEP(Y,Z|X), DEP(X,Z|Y).

Theorem 4.1 follows from our previous results and the results in Zhang and Spirtes (2008) as follows. If ($\mathcal{V}$,$\mathcal{E}$,P) is *adjacency faithful* in the sense of (2008, 253) and its empirical submodel satisfies condition (4.1)(a) of theorem 4, X,Y,Z must form an unshielded triple X–Y–Z in $\mathcal{E}$.

Then condition (4.1)(b) constitutes a violation of Zhang and Spirtes' *orientation faithfulness* (ibid.): Y must occur either in every or in no subset that screens off X from Z. The simplest example of (4.1) is given by a submodel ({X,Y,Z},P) with the (in)dependencies DEP(X,Y), DEP(X,Y|Z), DEP(Y,Z), DEP(Y,Z|X), but INDEP(X,Z) and INDEP(X,Z|Y): Here Y is con-

tained in the screen-off set {Y}, but not in the screen-off set ∅. An empirical model verifying (4.2) can be produced by canceling parents as in fig. 5 (cf. Zhang and Spirtes 2008, 256, fig. 6).

Since unfaithful causal systems exist, (C)+(F) would be empirically false if (F) were formulated as a strictly general axiom. Proponents of TCN argue that (F) is highly probable, i.e. satisfied by almost all empirical models. These arguments are based on the fact that unfaithful CMs are *parameter instable* in the following sense: their unfaithful independencies can be destroyed by arbitrary small changes of their *parameters* P(X|par(X)) without violating (C):

Lemma 2: A (C)-satisfying acyclic CM ($\mathcal{V}$,$\mathcal{E}$,P) is faithful *iff* it is parameter-stable.

Lemma 2 is well-known (Pearl 2009, 48, def. 2.4.1). It implies that for every probability measure over the set of parameters of a CM which is "smooth" (i.e. absolutely continuous with the Lebesgue measure over $[0,1]^p$), the probability of unfaithfulness is *zero* (cf. SGS 2000, 41f; Steel 2006, 313). This is a *formal* result on probabilities. It can only have implications for the frequency of unfaithful models in the *real* world if we assume that the parameters of causal models *do indeed vary* in the real world. But *can* these parameters always be varied? The belief that parameters can be varied is usually supported by the argument that the variables of causal systems are constantly perturbed by tiny influences from mutually independent noise variables in almost all real world domains (cf. Steel 2006, 313). On closer inspection, however, the argument faces two problems that require further investigation:

(1.) According to the standard assumptions of TCN, the parameters of a CM represent "autonomous mechanisms" (cf. Pearl 2009, 63). What is directly manipulable in a causal system without destroying its causal structure are not the parameters themselves, but the probability dis-

tributions over the variables by means of causal links to *external noise*. This noise is usually represented by "noise variables" which summarize the influence of all external perturbations on a variable (cf. Pearl 2009, 27; SGS 2000, 28). For the unfaithfulness triangle in fig. 3(b), this is graphically illustrated in fig. 6:

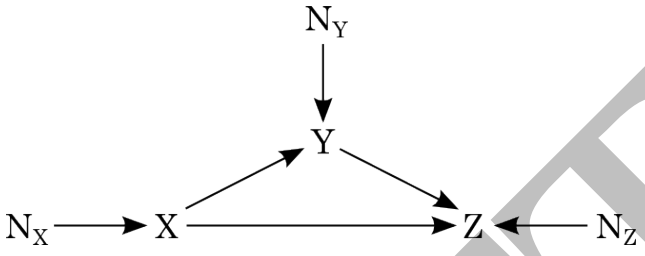


Fig. 6: Unfaithfulness triangle of fig. 3(b) with external noise variables N_X,N_Y,N_Z

By independently varying the prior probability distribution over the noise variables, the model's parameters can be changed according to the following equation, where P(n) = P(n|par(X)) because the noise variables are assumed to be mutually independent:

(14) $P(X|par(X)) = \Sigma_{n \in Val(N_X)} P(X|par(X),n) \cdot P(n)$.

Assuming that the external noise variables have probabilistic influence on X (i.e. $P(X|par(X),n) \neq P(X|par(X))$ for some n), we can change the parameter P(X|par(X)) by varying the prior probability $P(N_X)$. Given we do this independently for all noise variables, any unfaithfulness due to cancelation will disappear with high probability.

(2.) We must distinguish between *external* noise (caused by external perturbation variables) and *error* noise (caused by unrepresentative samples). External noise is a property of the population, while error noise is a property of the sample which decreases with increasing sample size N in proportion to $1/\sqrt{N}$. External noise will turn an unfaithful independence into a dependence, but it cannot turn a faithful independence into a dependence, since the noise influences are mutually independent in the population. Sampling errors, on the other hand, can also turn a faithful

independence in the population into an accidental dependence in the sample. Zhang and Spirtes (2003) show that the probability of an α-error (rejection of a true independence hypothesis) and a β-error (rejection of a true dependence hypothesis) can be held simultaneously small only if the parametric correlations in the population, P(X|par(X))−P(X), exceed a small threshold λ; Zhang and Spirtes (2003) call this stronger property *λ-faithfulness*. Uhler et al. (2013) show that for not too small λ, the probability p_λ that a CM violates λ-strong faithfulness may be quite high. We cannot offer a better solution to this problem than the remark that if the sample size N is very high, λ can be chosen so small that p_λ will still be low.

In this section we focus on (in)dependencies as population properties, by whose means TCN's content is expressed. Based on the preceding consideration we stipulate the following assumption about external noise:

(15) *External noise assumption (EN):* The variables of most causal structures in our world are causally influenced by many small and mutually independent disturbances (external noise) that fluctuate randomly over time.

The existence of external noise makes unfaithfulness due to cancelation highly improbable. But what about the other two kinds of unfaithfulness?

In a deterministic universe, (EN) does not render determinism unfaithfulness improbable, because the non-accidental deterministic dependencies "P(X|**Y**) = 1" that hold in such a universe hold by laws of nature: they cannot be "varied" because there are no noise variables which are not already included in the antecedent-set **Y**. To be sure, if determinism *and* the noise assumption are true, the number of causal parents that determine the value of a variable will be very high, but this doesn't change the force of the argument. Since we do not know whether the universe is deterministic, nor how probable this is, we cannot show that (EN) renders determinism

unfaithfulness improbable.

An even stronger skeptical argument applies to intransitivity unfaithfulness. Recall our example of an intransitive chain $X \rightarrow Y \rightarrow Z$ with $Val(Y) = \{y_1,y_2,y_3,y_4\}$ such that $INDEP(y_i|X)$ *iff* $i \in \{3,4\}$ and $INDEP(Z|y_i)$ *iff* $i \in \{1,2\}$). We assume a situation in which it follows from the nature of the *causal mechanisms* underlying $X \rightarrow Y$ and $Y \rightarrow Z$ that X has no causal influence on y_3 and y_4, and that neither y_1 nor y_2 has an influence on Z. In this case $INDEP(y_3,X|n_Y)$ and $INDEP(y_4,X|n_Y)$ will hold for all value-adjustments of the external noise variable N_Y (and the same goes for $INDEP(Z,y_i|n_Z)$) for $i \in \{1,2\}$. By equation (14) above, this implies that

(16) For $i \in \{3,4\}$: $P(y_i|X) = \Sigma_{n \in Val(N_Y)} P(y_i|X,n) \cdot P(n) = \Sigma_{n \in Val(N_Y)} P(y_i|n) \cdot P(n) = P(y_i)$

holds for *every* prior distribution P(N). So $INDEP(y_{i=3,4}|X)$ *cannot* be changed by adding external noise, and the same goes for $INDEP(Z|y_{i=1,2})$.

A frequently discussed example of this sort has been given by McDermott (1995): a right-handed terrorist is going to press a detonation button to explode a building, when a dog bites his right hand and causes him to use his left hand for pressing the button. McDermott's example is somewhat artificial, but there are more realistic examples from the sciences: Beryllium is diamagnetic in its ground state (magnetic moment m = 0), but paramagnetic in its first excited state, in which case its magnetic moment M takes one of several non-zero values m_i, depending on the direction of a given external magnetic field F. Assuming a beam of Beryllium atoms in a magnetic field, they will be deflected from the straight line L ($Val(L) = \{d,\neg d\}$) iff their magnetic moment is non-zero. Thus $Val(M) = \{0,m_1,\ldots,m_n\}$ and only the values $M = m_i$, but not the value $M = 0$, depend probabilistically on F, while D's value depends only on the value $M = 0$ vs. $M \neq 0$. So $F \rightarrow M \rightarrow D$ forms an intransitive causal chain in which F has no influence on D.

In conclusion, an empirically tenable version of a faithfulness axiom must be restricted to cancelation unfaithfulness. Cartwright (1999, 118) and Hoover (2001, 171) have objected that cancelation unfaithfulness is frequent in domains of self-regulatory systems whose parameter values have been selected by evolutionary or intentional processes. Steel (2006, 313) counters this objection by the argument that such selection processes can produce a precise cancelation of influences only if external noise is absent. We find this argument convincing and add that self-regulatory processes are never perfect and small deviations from a precise cancelation of influences occur.

In contrast, neither determinism nor intransitivity unfaithfulness is made improbable by external noise. Fortunately there is a solution of this problem: there exist purely empirical (or non-theoretical) conditions that are sufficient for the absence of determinism unfaithfulness and intransitivity unfaithfulness. They are stated in conditions (17.1+2) below. Our preliminary proposal of a tenable axiom of faithfulness asserts that if determinism and intransitivity unfaithfulness are excluded, causal models are with high probability faithful:

(17) *Axiom of restricted faithfulness (RF):* If a CM ($\mathcal{V}$,$\mathcal{E}$,P) satisfies (P) and its non-theoretical submodel ($\mathcal{V}$,P) satisfies conditions (17.1) and (17.2), then ($\mathcal{V}$,$\mathcal{E}$,P) is faithful with very high probability:

(17.1) (Exclusion of determinism unfaithfulness): No value of a variable X$\in$ $\mathcal{V}$ depends deter-

ministically on a subset $\mathbf{U} \subseteq \mathcal{V}-\{X\}$, and

(17.2) (Exclusion of intransitivity unfaithfulness): there exists no sequence of variables $(Z_1,\ldots,Z_n)$ such that

(a) $DEP(Z_i,Z_{i+1}|\mathbf{U})$ holds for all $\mathbf{U} \subseteq \mathcal{V}-\{Z_i,Z_{i+1}\}$ $(1 \leq i < n-1)$, but

(b) $INDEP(Z_1,Z_n|\mathbf{E})$ holds for $\mathbf{E} = \{Z_i: INDEP(Z_{i-1},Z_{i+1}), 1 \leq i \leq n\}$.

Assuming adjacency-faithfulness, condition (17.2)(a) implies that the variables form an undirected chain $X_1-X_2-\ldots-X_n$. So condition (17.2)(a)+(b) requires dependence-transitivity for every such chain conditional on a set **E** that contains all common effects on this chain.[16] Axiom (RF) asserts faithfulness on the condition that determinism and intransitivity unfaithfulness are excluded. It is justified by the external noise assumption, which makes all sorts of cancelation unfaithfulness highly improbable.

We understand axiom (RF) as a first suggestion; improvements of it are left to future papers. For example, condition (17.2) does not only exclude intransitivity unfaithfulness: it also excludes some cases of cancelation unfaithfulness (viz., those that are distinct from adjacency unfaithfulness). We couldn't avoid this disadvantage, since in order to make axiom (RF) empirically contentful, condition (17) had to be formulated in an empirical (non-theoretical) way. This empirical content is expressed in the following corollary of theorem 4:

---

[16] If $INDEP(Z_{i-1},Z_{i+1})$ holds, then either Z_i is a common effect and must be in **E** (the "intended case"), or Z_i is a common or intermediate cause which violates dependence-transitivity (in which case the inclusion of Z_i in **E** does no harm), or the dependence between Z_{i-1} and Z_{i+1} is canceled by a compensating path, which is made improbable by assumption (EN).

Corollary 1: Axioms (C)+(P)+(RF) imply that empirical (non-theoretical) models ($\mathcal{V}$,P) which satisfy conditions (17.1) and (17.2) of axiom (RF) and either condition (4.1) or (4.2) of theorem 4 are very improbable.

In conclusion, (RF) produces probabilistic empirical content when added to axioms (C)+(P): (C)+(P)+(RF) make empirical (or non-theoretical) models ($\mathcal{V}$,P) as described in corollary 1 very improbable. Moreover, this result holds for all possible variables, and hence, also for variables whose probability distribution has not yet been observed. So this empirical content generates novel predictions that are independently testable. The same is true for the empirical consequences of the stronger versions of TCN introduced in sec. 3.3 and 4.

It comes without surprise that if we measure the content of the empirical models excluded by axioms (C)+(RF) by a logical information measure based on an indifferent prior over the parameter space, then this content is very small – simply because the probability of the excluded empirical models is very small. Recall our comparison of "causality" in TCN with "force" in Newtonian mechanics. Obviously, (RF) doesn't play an analogous role as special force laws do in Newtonian mechanics, since (RF) increases TCN's content only a little bit. In the next subsection we show how TCN's content can be increased in a similar way as special force laws increase Newtonian mechanic's content, by introducing more "substantial" axioms that constrain the mechanisms underlying a CM's causal arrows.

3.3 Empirical content of temporal forward-directedness

If we add the assumption of temporal forward-directedness (T) of causal processes to TCN, the content of the resulting TCN-version is substantially increased. If (T) is joined with the conjunction of (C) and the faithfulness condition (F), it excludes the possibility that two variables are screened off by future common causes or linked up by a common effect lying in the past of one of the two. The possible and impossible situations are illustrated in fig. 7.

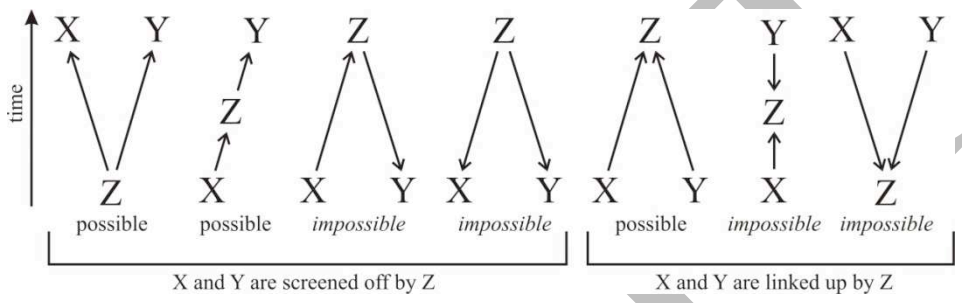


Fig. 7 Implications of temporal forward-directedness for screening off and linking up; X and Y can swap places in each structure

To make this idea precise, we define a *causal event-model* ($\mathcal{V}$,$\mathcal{E}$,P,t) as a CM ($\mathcal{V}$,$\mathcal{E}$,P) whose variables are event-variables, together with a time function t: $\mathcal{V}\rightarrow$Reals, where t(X) is the time point at which the possible X-values (events) x occur. We explicate axiom (T) as follows:

(18) *Axiom of temporal forward-directedness (T):* Every physically possible causal event model ($\mathcal{V}$,$\mathcal{E}$,P,t) [in an intended domain] satisfies *condition* (T), which is defined as follows:

$X \rightarrow Y$ implies $t(X) < t(Y)$.

Note that all physically possible causal event models satisfying (T) are acyclic. The strict generality of axiom (T) is an open question in contemporary physics,[17] whence we insert the cautious phrase in square brackets. The next theorem tells us that a variety of empirical models are excluded when (T) is added to (C)+(F). Screening off by future events is generally impossible; linking up by semi-past events (in the past of at least one linked up event) is only impossible when the linking up event and the linked up events are not screened off by common causes lying in their past. Fig. 8 presents an example where this condition is violated: here we have INDEP(X,Z) and DEP(X,Z|Y), although $t(Y) < t(X),t(Z)$.

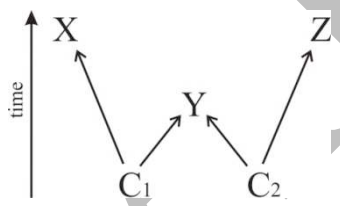


Fig. 8 X and Z are linked up by Y, though Y lies in the past of X and Z.

Theorem 5: (C)+(F)+(T) entail that empirical (non-theoretical) event-models ($\mathcal{V}$,P,t) with probabilistic dependencies of the following sort are impossible:

(5.1) *Screening off by future events:* DEP(X,Z), and INDEP(X,Z|Y), where $t(Y) > t(X)$ and $t(Y) > t(Z)$.

(5.2) *Linking up by semi-past events without past common causes:* INDEP(X,Z) and DEP(X,Z|Y), where the following holds for $\psi = X$ or for $\psi = Z$: (*) $t(Y) < t(\psi)$ and there ex-

[17] A universe in a high entropy state could admit temporally inverted causal processes (cf. Reichenbach 1956, 136ff; Savitt 1996, 353).

ists no C with t(C) < t(Y) and INDEP(ψ,Y|C).

Theorem 5 assumes the full faithfulness condition (F), which is not an axiom of TCN. However, we can also formulate a probabilistic version of theorem 5 that follows from (C)+(RF) and says that CMs of the sort (5.1) and (5.2) are highly improbable.

Reichenbach (1956, 162) did not assume faithfulness in his attempt to justify the direction of time on the basis of directed causal relations. He rather argued for the impossibility of a "fork open towards the past", i.e. a future event Z that screens off two simultaneous events X and Y without having a common cause Z' in their past. Theorem 6.1 proves Reichenbach's argument within TCN. Theorem 6.2 goes beyond Reichenbach's claim and shows that under the additional assumption of (RF), a future event Z can screen off X from Y only if either X, Y, or some values of a past screen-off set for {X,Y} depend deterministically on Z, where **U** is called a *past screen-off set* for X and Y iff INDEP(X,Y|**U**) and t(Z) < t(X),t(Y) for all Z∈**U**.

Theorem 6: Assume X,Y∈ $\mathcal{V}$ are two temporally simultaneous event-variables with DEP(X,Y).

Then (C)+(T) entail that no non-theoretical model ($\mathcal{V}$,P,t) can verify condition (6.1), and

(C)+(RF)+(T) entail that a non-theoretical model ($\mathcal{V}$,P,t) that verifies condition (6.2) is very

improbable:

(6.1) There exists no past screen-off set **U** for X and Y.

(6.2) There exists a variable Z in the future of X and Y screening off X from Y, and there exists a Z-value z on which no value of X, of Y, or of some past screen-off set **U** for X and Y depends deterministically.

Like theorem 5.2, theorem 6 refers to the (non-)existence of not necessarily observable variables; so its content is, strictly speaking, non-theoretical but not empirical.

In conclusion, axiom (T) strongly enriches the empirical and non-theoretical consequences of TCN. A still stronger enrichment is possible by assuming the condition of *locality*, which asserts that no causal influence in an event model is propagated with a speed greater than light velocity. A precise explication of this condition and its content is left to future work.

4. Conclusion

In this paper we investigated TCN in regard to its explanatory warrant and empirical content. Concerning the explanatory warrant we demonstrated that TCN's core axioms can be justified by an IBE of screening off and linking up (sec. 2). This justification leaves it open whether the core axioms of TCN are strictly general, i.e. hold for all domains. In sec. 3 we saw that TCN's core, i.e. axioms (C)+(P), is empirically empty. If restricted faithfulness (RF) and the external noise assumption (EN) are added, the resulting extended TCN-version acquires weak probabilistic empirical content. By adding more "substantial" principles such as the principle of temporal forward-directedness (T), TCN's content becomes strong.

We mention one further kind of "substantial" principle which concerns human interventions:

(19) *Independence of human interventions (HI):* Most of a person's actions I = i manipulating

variables of a person-external causal system ($\mathcal{V}$,$\mathcal{E}$,P) that are experienced as "free" are probabilistically independent of those variables in $\mathcal{V}$ that are non-effects of I.

Axioms (C)+(HI) make it highly probable that human interventions are value realizations of intervention variables I in the sense of SGS (2000, sec. 3.7.2), Eberhardt and Scheines (2007), or Woodward (2003, 98). Adding intervention variables to an empirical model excludes further probability distributions. Elaborating this idea is left to future work.

Funding: This work was supported by Deutsche Forschungsgemeinschaft, research unit "Causation | Laws | Dispositions | Explanation" (FOR 1063).

Acknowledgements: For important discussions we are indebted to Michael Baumgartner, Mathias Frisch, Clark Glymour, Andreas Huettemann, Marie I. Kaiser, Kevin Kelly, Jeff Ketland, Theo Kuipers, Hannes Leitgeb, Margaret Morrison, Paul Naeger, Stathis Psillos, Henk de Regt, Oliver Scholz, Markus Schrenk, Peter Spirtes, Jon Williamson, and two anonymous reviewers.

References

Armstrong, D. M. (1983): *What Is a Law of Nature?* Cambridge: Cambridge University Press.

Balzer, W., Moulines, C. U., and Sneed, J. D. (1987): *An Architectonic for Science*. Dordrecht: Reidel.

Beebee, H., Hitchcock, C., and Menzies, P. (Eds., 2009): *The Oxford Handbook of Causation*. Oxford : Oxford University Press.

Blalock, H. (1961): Correlation and Causality: The Multivariate Case, *Social Forces, 39*, 246-

251.

Carnap, R. (1956): The Methodological Character of Theoretical Concepts. In: H. Feigl and M. Scriven (Eds.), *The Foundations of Science* (pp. 38-76). Minneapolis: University of Minnesota Press.

Carnap, R. (1971): A Basic System of Inductive Logic, Part I. In: R. Carnap and R. Jeffrey (Eds.), *Studies in Inductive Logic and Probability* (pp. 33-166). Berkeley: University of California Press.

Cartwright, N. (1999): *The Dappled World: A Study of the Boundaries of Science*. Cambridge: Cambridge University Press.

Cartwright, N. (2007): *Hunting Causes and Using Them*. Cambridge: Cambridge University Press.

Eberhardt, F., and Scheines, R. (2007): Interventions and Causal Inference, *Philosophy of Science, 74*, 981-995.

Fales, E. (1990): *Causation and Universals*. London: Routledge.

French, S. (2008): The Structure of Theories. In: S. Psillos and M. Curd (Eds.), *The Routledge Companion to Philosophy of Science* (pp. 269-280). London: Routledge.

Friedman, M. (1974): Explanation and Scientific Understanding. *Journal of Philosophy, 71*, 5-19.

Glymour, C. (2004): Critical Notice. *British Journal for the Philosophy of Science, 55*, 779-790.

Hausman, D. (1998): *Causal Asymmetries*. Cambridge: Cambridge University Press.

Healey, R. (2009): Causation in Quantum Mechanics. In: Beebee et al. (Eds.), *The Oxford Handbook of Causation* (pp. 673-686). Oxford : Oxford University Press.

Hitchcock, C. (2010): Probabilistic Causation. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Winter 2011 Edition), URL = <http://plato.stanford.edu/archives/win2011/entries/causation-probabilistic/>.

Hoover, K. (2001): *Causality in Macroeconomics*. Cambridge: Cambridge University Press.

Kitcher, P. (1989): Explanatory Unification and the Causal Structure of the World. In P. Kitcher, and W. Salmon (Eds.), *Scientific Explanation* (pp. 410-505), Minneapolis: University of Minnesota Press.

Lauritzen, S. L., Dawid, A. P., Larsen, B. N., and Leimer, H.-G. (1990): Independence Properties of Directed Markov-Fields. *Networks, 20*, 491-505.

Lewis, D. (1970): How to Define Theoretical Terms. *Journal of Philosophy, 67*, 427-446.

McDermott, M. (1995): Redundant Causation. *British Journal for the Philosophy of Science, 40*, 523-544.

Norton, J. D. (2009): Is There An Independent Principle of Causality In Physics? *British Journal for the Philosphy of Science, 60*, 475-486.

Papineau, D. (1992): Can We Reduce Causal Direction to Probabilities? *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, Volume 1992, Volume Two: Symposia and Invited Papers*, 238-252.

Papineau, D. (1996): Theory-dependent Terms. *Philosophy of Science, 63*, 1-20.

Pearl, J. (1988, 1997): *Probabilistic Reasoning in Intelligent Systems*. San Francisco: Morgan Kaufmann.

Pearl, J. (2000, 2009): *Causality*. Cambridge: Cambridge University Press.

Psillos, S. (2009): Regularity Theories. In Beebee et al. (Eds.), *The Oxford Handbook of Causation* (pp. 131-157). Oxford: Oxford University Press.

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

Savitt, S. F. (1996): The Direction of Time. *British Journal for the Philosophy of Science, 47*, 347-370.

Sneed, J. D. (1971): *The Logical Structure of Mathematical Physics*. Dordrecht: Reidel.

Spirtes, P., Glymour, C., and Scheines, R. (1993, 2000): *Causation, Prediction, and Search*. Cambridge: MIT Press.

Steel, D. (2006): Homogeneity, Selection, and the Faithfulness Condition. *Minds and Machines, 16*, 303-317.

Suppes (1970): *A probabilistic Theory of Causality*. Amsterdam: North-Holland.

Tomasello, M. (1999): *The Cultural Origins of Human Cognition*. Cambridge: Harvard University Press.

Uhler, C., Raskutti, G., Bühlmann, P., and Yu, B. (2013): Geometry of the Faithfulness Assumption in Causal Inference. *Annals of Statistics, 41*, 436-463.

Verma, T. S. (1986). Causal Networks: Semantics and Expressiveness. Technical Report R-65, Cognitive Systems Laboratory, University of California, Los Angeles.

Woodward, J. (2003): *Making Things Happen*. Oxford: Oxford University Press.

Wright, S. (1921): Correlation and Causation. *Journal of Agricultural Research, 20*, 557-585.

Zhang, J., and Spirtes, P. (2003): Strong Faithfulness and Uniform Consistency in Causal Inference. *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence* (pp. 632-639). San Francisco: Morgan Kaufmann.

Zhang, J., and Spirtes, P. (2008): Detection of Unfaithfulness and Robust Causal Inference. *Minds and Machines 18*, 239-271.

Zhang, J., and Spirtes, P. (2011): Intervention, Determinism, and the Causal Minimality Condi-

tion, *Synthese, 182*, 335-347.

Appendix: Proofs of lemmata and theorems

Proof of lemma 1:

For (1.1): Assume $(\mathcal{V},\mathcal{E})$ is acyclic and X→Y in $\mathcal{E}$. For reductio, assume π is a path in $\mathcal{E}-\{X\rightarrow Y\}$ that connects X with Y and is activated by par(Y)−{X}. So π must have the form X−−Z←Y. Thus π must carry at least one common effect Z*; otherwise π would have the form X←←Y and $(\mathcal{V},\mathcal{E})$ would be cyclic. But since Z*∉par(Y)−{X}, π is blocked by par(Y)−{X}, contradicting our assumption.

For (1.2): Assume **U** ⊃ par(Y)−{X}. Both par(Y)−{X} and **U** d-separate X from Y in $(\mathcal{V},\mathcal{E}-\{X\rightarrow Y\})$ (by assumption and lemma 1.1, respectively). Direction ⇐ is trivial. For direction ⇒ we have to prove that INDEP(X,Y|par(Y)−{X}) implies INDEP(X,Y|**U**). We proceed by induction on the number of elements in **U**−(par(Y)−{X}). Assume the claim has been proved for some **U** ⊇ par(Y)−{X} and let **U**' = **U**∪{Z}, where **U** and **U**' d-separate X from Y in $(\mathcal{V},\mathcal{E}-\{X\rightarrow Y\})$. We show that

(*) Either Z is d-separated from X by **U**∪{Y}, or Z is d-separated from Y by **U**∪{X}.

For reductio, assume (*) does not hold. We distinguish two cases. Case (A): Z is d-connected with X given **U**∪{Y}, and with Y given **U**∪{X}, by two paths π_X: Z−−X and π_Y: Z−−Y, respectively, which both do not carry X→Y. In this case, Z is d-connected with X and with Y by the respective paths given **U** alone. Let π be the concatenation of π_X and π_Y. If Z is a common effect on π, then **U**' would activate a new X-Y-connecting path, which is excluded, and if Z is not a common effect on π, then π would d-connect X and Y given **U**, which is excluded. So case (A) is impossible. The other possible case is (B): Z is d-connected with Y and with X only by paths π that contain X→Y as a subpath; these paths must either have the form (i) X→Y−−Z or (ii) Z−−X→Y, but not both (else we would have case (A)). But this is impossible since in case (i) X is d-separated from Z by **U**∪{Y} (since if Y←Y' is on π, then Y'∈**U**), and in case (ii) Z is d-separated from Y by **U**∪{X}.

Dawid's axioms of probabilistic independence (cf. Pearl 1988, 84) include the following (probabilistically valid) axioms:

Contraction: INDEP(X,Y|{Z}∪**U**)∧INDEP(X,Z|**U**) ⇒ INDEP(X,{Y,Z}|**U**).

Decomposition: INDEP(X,{Y,Z}|**U**) ⇒ INDEP(X,Y|**U**)∧INDEP(X,Z|**U**).

Weak union: INDEP(X,{Y,Z}|**U**) ⇒ INDEP(X,Y|{Z}∪**U**).

Assume by (*) that Z is d-separated from X by **U**∪{Y}. (The other possibility is that Z is d-separated from Y by **U**∪{X}; the proof proceeds in exactly the same way.) Since ($\mathcal{V}$,$\mathcal{E}$,P) satisfies (C), (a) INDEP(X,Z|**U**∪{Y}) follows. From the induction hypothesis INDEP(X,Y|**U**) and (a) INDEP(X,Z|**U**∪{Y}) we get (b) INDEP(X,{Y,Z}|**U**) by contraction, and from (b) we get INDEP(X,Y|**U**∪{Z}), i.e. INDEP(X,Y|**U**') by weak union. Q.E.D.

Proof of theorem 2:

Proof of (P) ⇒ (Min): Assume ($\mathcal{V}$,$\mathcal{E}$,P) is not minimal. So there exists an X→Y in $\mathcal{E}$ such that ($\mathcal{V}$,$\mathcal{E}'$,P) satisfies (C), where $\mathcal{E}'$:= $\mathcal{E}$–{X→Y}. Since par(Y)–{X} d-separates X from Y in ($\mathcal{V}$,$\mathcal{E}'$) (by lemma 1.1), INDEP(X,Y|par(Y)–{X}) holds because of (C). So (P) is violated.

Proof of (Min) ⇒ (P): Assume that ($\mathcal{V}$,$\mathcal{E}$,P) satisfies (Min), which means that there is no X,Y∈$\mathcal{V}$ with X→Y∈$\mathcal{E}$ such that ($\mathcal{V}$,$\mathcal{E}$–{X→Y},P) still satisfies (C). The latter is the case iff

(*) the parent set par(Y) of every Y∈$\mathcal{V}$ (with par(Y) ≠ ∅) is minimal in the sense that removing one of Y's parents X from par(Y) would make a difference for Y, meaning that P(y|x,par(Y)–{X}) ≠ P(y|par(Y)–{X}) holds for some X-value x, Y-value y, and some instantiations of par(Y)–{X}.

For otherwise P would admit the Markov factorization according to (8.2) both relative to ($\mathcal{V}$,$\mathcal{E}$,P) and relative to ($\mathcal{V}$,$\mathcal{E}$–{X→Y},P). This implies by theorem 1 that ($\mathcal{V}$,$\mathcal{E}$,P) and

($\mathcal{V}$,$\mathcal{E}$–{X→Y},P) satisfy (C), i.e. ($\mathcal{V}$,$\mathcal{E}$,P) is not minimal, which contradicts our assumption.

Now, (*) entails that Dep(X,Y|par(Y)–{X}) holds for all X,Y∈$\mathcal{V}$ with X→Y, i.e., that ($\mathcal{V}$,$\mathcal{E}$,P) satisfies (P). Q.E.D.

Proof of theorem 5:

For (5.1): Recall Dawid's axioms of probabilistic independence from the proof of lemma (1.2). By switching Y with Z and setting **U** = ∅, the contraposed forms of decomposition and contraction give us DEP(X,Z) ⇒ DEP(X,{Y,Z}) and DEP(X,{Y,Z}) ∧ INDEP(X,Z|Y) ⇒ DEP(X,Y). In the same way, switching X with Y and setting **U** = ∅ gives us DEP(Y,Z) ⇒ DEP(Y,{X,Z}) and DEP(Y,{X,Z}) ∧ INDEP(X,Y|Z) ⇒ DEP(Y,Z).

By these considerations, the assumptions DEP(X,Z) and INDEP(X,Z|Y) of theorem 5.1 entail DEP(X,Y) and DEP(Y,Z). So (by (C)) X and Y as well as Y and Z are d-connected given ∅ by two paths X––Y and Y––Z, whence (a) these two paths don't carry a common effect. (F) implies that X and Z are d-separated by Y which together with (a) implies (b) that X–...→Y←...–Z is impossible. (a)+(b) entail that either X←←Y or Y→→Z. But both possibilities are excluded by condition (T).

For (5.2): Because of (F) and INDEP(X,Z), X and Z are d-separated (by ∅). Thus and because of (C) and DEP(X,Z|Y), X and Z are d-connected by a path π that carries a common effect Y' that is either identical with Y or has Y as an effect. So π: X–…→Y'←…–Z, where π contains

no colliders except Y'. By condition (T) and assumption (*), this path cannot carry a common cause of Y' and X and one of Y' and Z. So either (a) X→→Y' or (b) Y'←←Z must be a subpath of π. Both cases are excluded: In case (a), t(Y) < t(X) holds by assumption (*), whence also t(Y') < t(X) must hold (since either Y'=Y or Y'→→Y which implies by (T) that t(Y') < t(Y)). The same argument applies to case (b). Q.E.D.

Proof of theorem 6:

For (6.1): By DEP(X,Y) and (C), X and Y are d-connected (by ∅), and by (T) and t(X) = t(Y), X and Y can only be d-connected by common causes Z lying in their past. By (C), conditionalization on the set **U** of all such common causes must screen off X from Y.

For (6.2): Assume Z is in the future of X and Y screening off X from Y, whence DEP(X,Y) and INDEP(X,Y|Z). It follows from the axioms of decomposition and contraction (similar as in the proof of theorem 5.1) that DEP(X,Z) and DEP(Y,Z) hold. Thus, X and Y, X and Z, and also Y and Z must be d-connected given ∅. From this together with (C), (T) and t(Z) > t(X) = t(Y) it follows that X and Y must be d-connected by a common cause path π_U: X←←**U**→→Y (where **U** is the set of all common causes of X and Y), and that X and Z must be d-connected by a cause-effect or a common cause path; the same holds for Y instead of X. So there will also be a path X–…→Z'←…–Y, where Z' is the only collider on this path and either (i) Z' = Z or (ii) Z'→→Z holds. *In what follows we write Z for Z'.*

Let **P** = par(X)∪par(Y) be the set of all parents of X and of Y. By the Markov-condition (M) and (T), **P** is a past screen-off set for {X,Y} (though a redundant one: par(X) or par(Y) alone is one, too); so INDEP(X,Y|**P**) holds. Note also that for some **p**, DEP(X,**p**) and DEP(Y,**p**) must hold, since otherwise, by the proof in footnote 10, the path π_U that d-connects X and Y could not transmit dependence between X and Y.

There are two possible cases: either (A) INDEP(X,Y|**P**∪{Z}) or (B) DEP(X,Y|**P**∪{Z}). As-

sume (B) is the case. Then we have INDEP(X,Y|Z) and DEP(X,Y|**P**∪{Z}), i.e. X and Y are linked up and thus d-connected given Z conditional on **P**, but not unconditionally. In other words, conditionalizing on **P** isolates the common effect path X –…→Z←…–Y between X and Y which was exactly canceled by π_U before. This would be a case of cancelation unfaithfulness, which is, according to (RF), highly improbable.

Now assume case (A), INDEP(X,Y|**P**∪{Z}). Since X and Y are d-connected given **P**∪{Z}**,** this constitutes, again, a case of unfaithfulness. We will show that it is (a) either one of cancelation and, hence, made improbable by axiom (RF), or (b) a case of deterministic dependence of some X- or **P**-value on every z∈Val(Z). Let us assume that (b) is false, i.e.:

(*) ∀**p**∀x∃z: P(**p**,x,z) ≠ 0,1.

Without restricting the assumption of our proof we can assume that the prior probabilities of all values of **P**, X, and Z are positive; simply by removing all values with zero-probability from the value-space.

We compute, attaching indices to the identity signs for easy reference:

$$\begin{array}{ccccc} P(Y|X,Z) & =_1 & \Sigma_{\mathbf{p}} P(Y|\mathbf{p},X,Z)\cdot P(\mathbf{p}|X,Z) & =_2 & \Sigma_{\mathbf{p}} P(Y|\mathbf{p},Z)\cdot P(\mathbf{p}|X,Z) \\ \vdots & & & & \vdots \\ =_3 & & & & =_4 \\ \vdots & & & & \vdots \\ P(Y|Z) & =_5 & & & \Sigma_{\mathbf{p}} P(Y|\mathbf{p},Z)\cdot P(\mathbf{p}|Z) \end{array}$$

Identity "$=_3$" holds by assumption INDEP(X,Y|Z), given (*). (Note that by the definition of "INDEP(X,Y|Z)" ∀x,y,z: P(y|x,z) = P(y|z) or P(x|z) = 0 or P(z) = 0 holds; the latter case is excluded by our assumption of positive prior probabilities.) The identities "$=_1$" and "$=_5$" hold by probability theory, and "$=_2$" follows from case (A) (plus assumption (*) and positive priors). This gives identity "$=_4$". The identity "$=_3$" can only hold in two cases:

Case (A.1): DEP(Y,**P**|Z), i.e. there exist at least two distinct values P(Y|$\mathbf{p}_1$,Z) ≠ P(Y|$\mathbf{p}_2$,Z) in the above sums at the right hand side. Assume P(**p**|X,Z) differs from P(**p**|Z) for some **P**-values **p**.

The values of P(**p**|X,Z) and P(**p**|Z) are weights whose sums always add up to one. Since the differences in these weights do not change the resulting sums ($\Sigma_{\mathbf{p}}$P(Y|**p**,Z) · weight$_{\mathbf{p}}$), this can only be because these differences are exactly canceled by the differences in the P(Y|**p**,Z)-values. This would constitute a case of unfaithfulness due to internal canceling paths (in the sense of Naeger, see sec. 3.3), which is made improbable by axiom (RF). So we infer P(**p**|X,Z) = P(**p**|Z) for all **P**-values **p**, i.e. INDEP(X,**P**|Z).

Case (A.2): INDEP(Y,**P**|Z). From the two cases (A.1+2) we conclude:

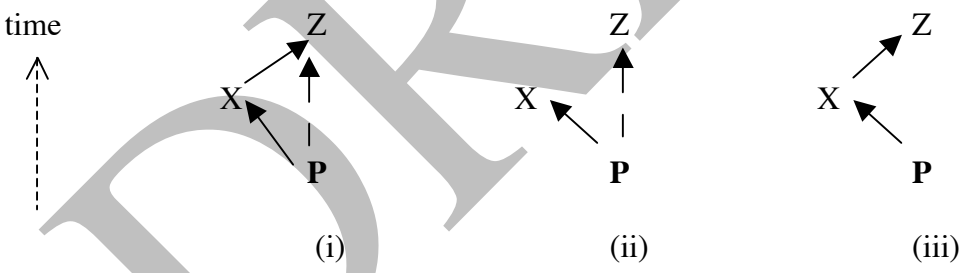
Case (A):* INDEP(X,**P**|Z) ∨ INDEP(Y,**P**|Z), i.e. Z screens off either X or Y from **P**.

For the rest of the proof we assume

(a) INDEP(X,**P**|Z).

If INDEP(Y,**P**|Z), X and Y change their roles and the proof works in exactly the same way.

By the remarks above, the causal structure between X, **P** and Z has one of the following three forms:



Note that the path(s) from **P** to Z in (i) and (ii) may or may not include a common cause path (that may go through Y); this is indicated by "– →".

Assume causal structure (i): We assume (+) DEP(X,Z|**P**) – otherwise the case is treated exactly as in the proof for structure (ii), which rests on condition INDEP(X,Z|**P**). Likewise we assume (++) DEP(**P**,Z|X) – otherwise the case is treated exactly as in the proof for structure (iii), which rests on INDEP(**P**,Z|X).

From (+) and (++) it follows that DEP(Z,**P**) must hold (otherwise, INDEP(**P**,Z) and IN-

DEP(**P**,X|Z) would imply by the axioms of contraction and decomposition INDEP(**P**,X), which contradicts DEP(**P**,X)). This means that we have a case of cancelation unfaithfulness: though both paths **P**→X and **P**–...→Z←X transmit probabilistic **P**-X-dependence when conditionalizing on Z, they exactly compensate each other. This case is made improbable by axiom (RF).

Assume causal structure (ii): Here we have INDEP(X,Z|**P**) by the causal Markov condition (M).[18] From INDEP(X,Z|**P**), INDEP(X,**P**|Z) and our assumption of positive priors we get:

(b) $\forall x,z,\mathbf{p}$: $P(x|z,\mathbf{p}) = P(x|z) \vee P(\mathbf{p}|z) = 0$, and

(c) $\forall x,z,\mathbf{p}$: $P(x|z,\mathbf{p}) = P(x|\mathbf{p}) \vee P(\mathbf{p}|z) = 0$.

(d) By DEP(X,**P**) exist $\mathbf{p}_1,\mathbf{p}_2$ such that $P(x|\mathbf{p}_1) \neq P(x|\mathbf{p}_2)$.

Thus either $\forall z$: $P(\mathbf{p}_1|z) = 0 \vee P(\mathbf{p}_2|z) = 0$, or for some z: $P(x|\mathbf{p}_i) = P(x|z)$ holds by (b)+(c) for i = 1 and i = 2, which contradicts (d). So $\forall z$: $P(\mathbf{p}_1|z) = 0 \vee P(\mathbf{p}_2|z) = 0$), i.e. every Z-value has some **P**-value that depends deterministically on it.

Finally assume causal structure (iii): Here we have INDEP(**P**,Z|X) since **P** and Z are d-separated by X. INDEP(**P**,Z|X) and INDEP(**P**,X|Z) plus positive priors give us (similarly as in case (ii))

(b') $\forall x,z,\mathbf{p}$: $P(\mathbf{p}|z,x) = P(\mathbf{p}|x) \vee P(x|z) = 0$, and

(c') $\forall x,z,\mathbf{p}$: $P(\mathbf{p}|z,x) = P(\mathbf{p}|z) \vee P(x|z) = 0$.

(d') By DEP(X,**P**) exist x_1,x_2 such that $P(\mathbf{p}|x_1) \neq P(\mathbf{p}|x_2)$.

Thus either $\forall z$: $P(x_1|z) = 0 \vee P(x_2|z) = 0$ or for some z: $P(\mathbf{p}|x_i) = P(\mathbf{p}|z)$ holds by (b')+(c') for i = 1 and i = 2, which contradicts (d'). So every Z-value has some X-value that depends deterministically on it.

Thus, assumption (*) must be false, which concludes our proof. Q.E.D.

---

[18] What we prove is slightly stronger than what follows from Dawid's axiom of intersection (Pearl 1988, 84) applied to INDEP(**P**,Z|X) and INDEP(**P**,X|Z).