Causality and Causation in DOLCE

Jos LEHMANN, Stefano BORGO, Claudio MASOLO, Aldo GANGEMI
Laboratory for Applied Ontology, ISTC-CNR, Trento and Rome, Italy
{*lehmann, borgo, masolo, gangemi*}*@loa-cnr.it, http://www.loa-cnr.it*

Abstract. We present a general conceptualization of causal relations that pivots on the distinction between *causality*, a law-like relation between types of events, and *causation*, the actual causal relation that holds between individual events. This distinction finds its formal characterization and embedding within DOLCE, in terms of a number of dependences between (types of) *quality changes*. Finally an application of the presented theory to the classical example of "the broken window" is provided.[1]

1 Introduction

The definition of sufficiently general theories of causal relations is among the toughest problems faced by most philosophical and scientific communities. The bulk of the problem lies in defining a theory that captures enough causal phenomena without, though, making causal relations too pervasive. Finding this balance is necessary for generating correct, yet non redundant explanations and/or predictions. In this paper we present a framework for representing causal relations between (types of) events in DOLCE, the Descriptive Ontology for Linguistic and Cognitive Engineering defined in [1]. The general conceptualization of causal relations presented in sections 2 and 3 is based on a reworking of [2, 3], and it pivots on the distinction between *causality*, a law-like relation between types of events, and *causation*, the actual causal relation that holds between individual events. In section 4, this distinction finds its formal characterization and embedding within DOLCE, in terms of a number of dependences between (types of) *quality changes*. Section 5 we show how theory applies to the canonical example of this paper: the broken window. Finally, in section 6 we draw some conclusions.

2 Existing approaches and their limitations

The starting point of our presentation of existing approaches is the following rephrasing of an example deviced by Ducasse in [4].

Example 1 (The Broken Window). Whilst a canary sings in the immediate proximity of a window of a house, a brick is thrown at the glass of the window, which breaks down.

Readers intuition mostly seems to suggest that the question "What did cause the window glass to break down?" should be answered by indicating the brick and/or the event undergone by the brick as the cause. There are a number of approaches to the problem of defining the (general) theory of causal relations underlying causal assessments such as the one just given. In the following, we briefly discuss these approaches and highlight their limitations.

[1]Work partially supported by the EU projects IST-2001-33052 (WonderWeb), FP6-507164 (Metokis) and the research grant "Logical Instruments for Ontology Analysis" of the Provincia Autonoma di Trento.

First, we note that, depending on the approach, the main variables C (for cause) and E (for effect) range over different "things". In the case of logical approaches, the variables obviously range over propositions. It should be noted that different sets of propositions may intuitively be considered as specifying Example 1 in different degrees. For instance, we consider M_2 below to be more specific than M_1, in the sense that proposition a intuitively implies proposition α, b intuitively implies β and so on[2]. In the case of singularistic approaches, the main variables C and E are meant to range over untyped events. In this case, the members of M_1 are types of events while the members of M_2 are individual events. Analogously for probabilistic approaches, even though these approaches may also be used on propositions.

1. M_1 consists of: α = an object vibrates, β = an object moves, γ = an object collapses.

2. M_2 consists of: a = the canary sings, b = the brick is thrown, c = the glass breaks down.

Taking these as legitimate specifications, in what follows, we use them for highlighting our evaluation of existing approaches relative to the task of stating and justifying part of the preferred causal interpretion of Example 1. According to such interpretation, a causal relation should be inferred between β and γ in M_1 and between b and c in M_2.

Secondly, given the scope of Example 1, we clearly evaluate the considered approaches only relative to their adequacy in capturing causal relations of a *physical* nature. It is reasonable to expect, though, that the limitations pointed out in the following hold also when the evaluated approaches are applied to the assessment of causal relations of a different nature, like *agentive* or *negative*.

Thirdly, reasoning about Example 1 at the level of M_1 can be seen as relevant to M_2 too: causal relations in M_1 are usually called *law-like* causal relations. A relation of this kind between members of M_1 can constrain the actual causal relations between the members of M_2. The converse does not hold, though: if, according to a theory defined on M_2, an actual causal relation holds between, say, b and c, then no law-like causal relation should be inferred. This is to say that reasoning about Example 1 at the level of abstraction of M_2 may be correct but irrelevant for (theories defined on) M_1.

Logical approaches. The traditional definition of causal relations is in terms of logical conditions [5]. According to logical reductionism, the sentence *C is the cause of E* is equivalent to either of the following logical statements: *C is a necessary condition of E* ($E \rightarrow C$); *C is a sufficient condition of E* ($C \rightarrow E$); *C is a necessary and sufficient condition of E* ($C \leftrightarrow E$); *C is an INUS condition of E* ($(C \wedge X) \vee Y \leftrightarrow E$, where X is the causal context of the cause, Y is a series of alternative sets of causes and causal contexts, and the acronym INUS means *Insufficient but Non redundant part of an Unnecessary but Sufficient* condition, [6]). Limitation: these approaches, which are suited for defining theories of law-like causal relations (at the level of M_1), fail to capture the intuitive distinction between simple causal conditions and causes, when applied to the level of spefication of M_2. The approach based on INUS conditions, for instance, that is geared to the solution of this problem, provides no clear indications of how to solve it in a principled way. For C and E ranging over M_2, given E=c, it is unclear how to choose between the two admissible cases: (i) C = a, X = b, (ii) C = b, X = a. In other words, no criterion is provided for distinguishing between causes and contexts.

[2] If the (formal) language is fixed, this intuition may be brought to its extreme and M_1 may be taken as a *minimal specification* of Example 1 and M_2 as a *maximal* spefication of Example 1, that is meant to more adherently model the *actual* and *unrepeatable* (causal) sequence described in Example 1.

Probabilistic approaches. According to probabilistic approaches the sentence *C is the cause of E* is equivalent to the statement: *the conditional probability of E given C is higher than the conditional probability of E given not C* ($p(E|C) > p(E|\neg C)$). Limitation: also this approach is more suited for defining theories of law-like causal relations (at the level of M_1). Still, for C and E ranging over M_1, it is extremely difficult to fill in a frequency table containing statistics about a significant number of cases that are truly comparable to each other. And even when this is possible, Probability Theory, similarly to Logic, provides no principled way for distinguishing between genuine causes and mere co-occurrences (a well known problem, see [7]). This limitation is even stronger in probabilistic approaches than it is in logical ones, as it concerns both levels (M_1 and M_2).

Singularistic approaches. According to singularism the sentence *C is the cause of E* is equivalent to either of the following statements: *C is a counterfactual condition of E* ($C \rightarrowtail E$) [8]; *the cause of a particular change E is such particular change C as alone occurred immediately before in the immediate environment of E* [4]; *C has transfered some energy or momentum to E* [9, 10, 11, 12]. Limitations: these approaches, which *by definition* are meant for theories of actual causal relations (at the level of M_2), fail in various ways. Counterfactuals fails in so-called cases of overdetermination, like, for instance, a modified version of Example 1, where, instead of a singing canary, there is a baseball moving at the same time of b, towards the window, with the same momentum (i.e., in M_2, instead of a take a^*: a^*= the baseball moves). Given E=c, neither of the intuitive conclusions C=a^* or C=b can be implied through counterfactual test. On the other hand, the notions of change, uniqueness, temporal and spatial immediacy, energy need to be further constrained in order to capture actual causal relations.

3 The contribution of Formal Ontology

None of these approaches presents a deep logical/ontological analysis of the propositions or of the (types of) events over which C and E should range. In our view, formal ontology provides the tools for overcoming this impasse as it will become clear below. First, we informally illustrate some of the notions and philosophical assumptions behind our formalism.

According to the ontological framework we propose, causal relations relate events and event types, rather than (more or less specific) propositions. As often indicated in the literature, events have a strong causal flavour, due to their tight relationships with the notions of change and time, and this makes them intuitively appealing causal relata. Furthermore, in our vision, causal relations relate very simple (types of) events that change *one single* aspect of *one single* object. The main rationale behind such a restrictive definition of event is methodological. The formal characterization of causal relations between *simple* events presents two main advantages: firstly, it provides valuable insight on dealing with the intricate problem of defining causal relations between complex events; secondly, it helps in assessing the ontological correctness of models such as M_1 and M_2, because it forces the modeler to disentangle relevant ontological issues, while refining the model. For instance, when giving the ontological specification of α, β and γ, one is forced to make explicit and/or to choose between a series of ontological assumptions like: what changes in an object that vibrates? Its shape, its volume, its location? Which of these is primary, if any? If all of these aspects of the object change, which kind of relationships hold among them?

These are crucial questions that we believe must be answered to establish causal relationships between events and, thus, our first goal is to isolate types of dependence relations that one can embrace to answer these questions. We find three types of existential dependences that are formally defined in section 4.4: structural dependences, causality dependences and circumstantial dependences. Stating these dependences between sets of simple events allows us to introduce the causal constraints we need in order to express causation.

Structural constraints. These hold between a specific kind of *simple* events and a number of other kinds of *simple* events that *synchronically* change *the same* object with respect to different aspects. For instance, a structural constraint may impose that if an object changes shape, then it changes (simultaneously) its location as well. Analogously, one may constrain changes in mass and changes in volume of an object. Such constraints are structural in the sense that, based on the structure (thus, the ontological characteristics) of the objects, they define clusters of types that are *necessarily* associated with a given type of change. Notice that structural constraints do not depend on the degree of the considered change, they do not depend on measuring - we might even have no measure at all. No specific information on the event is needed beside the fact that the object is changed in the given aspect (say, shape), this alone allows us to infer that the object also changed in other aspect(s) (say, location).

Causality constraints. These hold between a specific kind of *simple* events and a number of other kinds of *simple* events that *diachronically* change *distinct* objects. The changes may be on different aspects of the objects. For instance, a causality constraint may impose a dependence between the change in location of an object (like in β) and a later (or earlier) change in shape of another object (like in γ). A similar dependence may be stated between changes in shape of an object (like in α) and changes in shape of another object (like in γ). We call these causality constraints because they state very general assertions between qualities of (different) objects taking into account the (admittedly weak) temporal relation between events. Even causality constraints define clusters of types of changes. By means of causality constraints one can say that, for instance, the movement of an object may be caused by the movement of another object and, at the same time, implicitly exclude that the change in shape of an object is caused by the change in color of another object. Similarly to structural constraints, causality constraints do not depend on the degree of the considered change: they do not depend on measuring - again, one might have no measure at all. It suffices to know that a change has modified the object with respect to, say, its location, to infer that another object has changed (or will change) in some aspect(s), like its location or its shape or both.

Circumstantial constraints. These bind event types taking into account the degree of the change they bring about as well as the types of objects they change. In this sense circumstantial costraints are very different from the types of constraints presented above; they depend on how refined our measures are - i.e. one needs to measure the changes to verify if these constraints are satisfied.

There are two main groups of circumstantial constraints:

1. *Intrinsic constraints* constrain two specific kinds of *simple* events to comply with restrictions either on the way the change is brought about or on the type of objects that are changed. For instance, in M_2, for a causal relation to hold between b and c, on the one hand, b should involve a translation and, on the other hand, c should be the change of a frangible glass;

2. *Relational constraints* constrain the relations between two specific kinds of *simple* events and/or the objects they change. For example, in M_2, for a causal relation to hold between b and c, b should temporally precede c and the location of the brick at the end of b should be the location of the glass at the beginning of c.

Causation. Causation is the relation that holds between two individual events that satisfy the causality and circumstantial constraints introduced on their types. Furthermore, the applications of our framework to the assessment of causation is sensitive to the adopted structural constraints, which filter out spurious causal relationships between a candidate effect and the events that are synchronically dependent on the candidate cause (see section 4.5).

4 The proposed formalization

In this section we formally introduce the distinctions between the three types of dependences/constraints that have been discussed above. Then, we use them to define causation.

4.1 Basic Notions from DOLCE

First, we informally present the predicates of DOLCE (the foundational ontology we take as underlying paradigm) used in our theory of causal relations (see [1] for the formal characterization of these predicates and more details on DOLCE).[3]

- $PED(x)$ stands for "x *is a physical endurant*", i.e., an entity located in space and time that is *wholly* present at any time it is present, e.g., a car, George Bush, the K2, an amount of gold, etc.
- $PD(x)$ stands for "x *is a perdurant/event*"[4], i.e., an entity that is only partially present at any time it is present, in the sense that some of its temporal parts may not be present, e.g., reaching the summit of the K2, a conference, eating, being open, etc. In particular we focus on perdurants with physical participants.
- $PQ(x), TQ(x)$ stand for "x *is a physical quality*", "x *is a temporal quality*", respectively. Qualities are basic 'aspects' of entities that can be perceived and measured like shapes, colors, lengths, speeds, energies. In this sense, they represent partial characterizations of an entity and depend existentially on it: every endurant (perdurant) comes with its physical (temporal) qualities, which exist as long as the endurant (perdurant) exists. In DOLCE the physical and temporal qualities are partitioned into a finite-set of *quality types*. We write $PQ_1, \ldots, PQ_n$ ($n \geq 1$) for the physical quality types and $TQ_1, \ldots, TQ_m$ ($m \geq 1$) for the temporal quality types. From (Ad50), in DOLCE, the spatial locations (SL) of physical endurants are qualities (therefore SL is one of the PQ_i), and from (Ad49), the temporal locations of perdurants (TL) are temporal qualities (therefore TL is one of the TQ_i).
- $PR(x), TR(x)$ stand for "x *is a physical region/quale*", "x *is a temporal region/quale*", respectively. Regions/qualia describe how a quality is 'classified' (positioned) within a specific space of regions called the *quality space*. A specific shade of red can be a quale for different color qualities inhering to different objects: the fact that two roses have the same color is represented by introducing two distinct color qualities (one for each rose), that have (at a certain time) the same position in the quality space of color, that is, at that time they have the same *color quale*. Quality types are in a one-to-one correspondence with quality spaces, therefore let us write $PR_1, \ldots, PR_n$ for the *physical quality spaces* corresponding to $PQ_1, \ldots, PQ_n$ and $TR_1, \ldots, TR_m$ for the *temporal quality spaces* corresponding to $TQ_1, \ldots, TQ_m$. In particular, *space region* (S) corresponds to *spatial location* and *temporal interval* (T) to *temporal location* (see [1]).

[3] When referring to DOLCE's axioms and definitions, we use the notation introduced in [1], where (Dd#) indicates a definition and (Ad#) an axiom.

[4] In this paper, the terms 'perdurant' and 'event' are used as synomyn.

- $\mathsf{PP}(x,y)$ stands for "*x is a proper part of y*" for perdurants and physical/temporal regions (see (Dd14) in [1]).

- $\mathsf{PC}(x,y,t)$, $\mathsf{PC_C}(x,y)$ respectively stand for "*the endurant x participates to the perdurant y during the time t*", "*the endurant x participates to the perdurant y during the whole duration of y*" (see (Dd63) in [1]).

- $\mathsf{IN}_T(x,y)$, $\mathsf{CN}_T(x,y)$ respectively stand for "*x is temporally included in y*" (see (Dd42) in [1]), "*x is temporally coincident with y*" (see (Dd48) in [1]).

- $\mathsf{qt}(q,x)$ stands for "*q is a quality of x*". With this relation we can say that the above physical and temporal quality types cover all the possible qualities of endurants and perdurants:

  (A1) $\mathsf{qt}(q,x) \rightarrow ((PED(x) \leftrightarrow (PQ_1(q) \vee \cdots \vee PQ_n(q))) \wedge$
  $(PD(x) \leftrightarrow (TQ_1(q) \vee \cdots \vee TQ_m(q))))$

- $\mathsf{ql}(r,q)$, $\mathsf{ql}(r,q,t)$ respectively stand for "*r is the quale/region of the perdurant's quality q*", "*r is the quale/region of the endurant's quality q during the time t*". The one-to-one correspondence between quality types and quality spaces is captured by these axioms:

  (A2) $\mathsf{ql}(r,q) \rightarrow (TQ_i(q) \leftrightarrow TR_i(r))$
  (A3) $\mathsf{ql}(r,q,t) \rightarrow (PQ_i(q) \leftrightarrow PR_i(r))$

  For notation conciseness we introduce the following relation stating that the quality of type PQ_i of x has quale r at time t:

  (D1) $\mathsf{ql}(PQ_i,r,x,t) \triangleq \exists q(\mathsf{qt}(q,x) \wedge PQ_i(q) \wedge \mathsf{ql}(r,q,t))$

4.2 Theory of Time

DOLCE is not committed to a particular theory of time. As this is necessary to express temporal constraints in causal relations, we adopt Allen and Hayes's theory [13], which is well known and expressive enough to capture the temporal constraints we need. However, our approach is not bound to this theory and it can be easily reformulated in terms other theories.

Allen and Hayes's theory formalizes a discrete and linear time using the binary primitive *meets* ($\|$) between convex and extended intervals. The intended interpretation of this primitive is: $t_1\|t_2$ if and only if the right extreme of the closure of t_1 coincides with the left extreme of the closure of t_2. Considering their temporal extensions, this relation is naturally extended to perdurants and to pairs formed by a time interval and a perdurant. On the basis of *meets*, the temporal precedence relation can be defined in the classical way:

(D2) $e_1 <_T e_2 \triangleq \exists t_1,t_2(t_1\|e_1 \wedge t_1\|t_2\|e_2)$.

4.3 Basic Quality Changes

Informally, a *basic quality change*[5] is a perdurant capturing the change of an endurant along just one aspect/quality type. Complex perdurants can be decomposed into quality changes by fixing: (i) one specific participant; (ii) one physical quality type along which the participant changes. Therefore, basic quality changes can be further decomposed only considering

[5]In the following we use the expressions "basic quality change" and "quality change" as synonyms.

shorter temporal extensions or proper parts of their participants. Consider, for example, the simultaneous change of the *color* and *shape* of an object x. This perdurant can be decomposed into two basic quality changes: the change of the color of x and the change of the shape of x. These quality changes can be further decomposed by considering, say, the first 3 seconds of the change of the shape of x or the change of the color of the right part of x, but they do not have proper parts with the same temporal extension and the same participant (the whole x).[6]

Following this informal description, given an endurant x and a physical quality type PQ_i, we define a basic quality change e in the following way:

(D3) $\mathsf{BQC}(e,x,PQ_i) \triangleq \mathsf{BQC^*}(e,x,PQ_i) \wedge \bigwedge_{j=1}^{n} \neg\exists e'(\mathsf{PP}(e',e) \wedge \mathsf{CN}_T(e',e) \wedge \mathsf{BQC^*}(e',x,PQ_j))$ (*Basic Quality Change*)

where:

(D4) $\mathsf{BQC}^*(e,x,PQ_i) \triangleq \mathsf{UPC_C}(x,e) \wedge \exists t,t',r,r'(\mathsf{IN}_T(t,e) \wedge \mathsf{IN}_T(t',e) \wedge \mathsf{ql}(PQ_i,r,x,t) \wedge \mathsf{ql}(PQ_i,r',x,t') \wedge r \neq r')$

(D5) $\mathsf{UPC_C}(x,e) \triangleq \mathsf{PC_C}(x,e) \wedge \forall y,t,t'((\mathsf{PC}(x,e,t) \wedge \mathsf{PC}(y,e,t')) \rightarrow x = y)$

For each quality type PQ_i, we define a predicate PQ_i^C (called *change type*) that individuates the set of basic quality changes with respect to PQ_i. These definitions are given by the following schema:

(D6) $PQ_i^C(e) \triangleq \exists x(\mathsf{BQC}(e,x,PQ_i))$

For example, given the quality type $Color$, $Color^C(e)$ holds if and only if e is a basic quality change with respect to the $Color$ quality type, that is, during e, the color quality of its (unique) participant x assumes different positions in the color space.

4.4 Quality Dependences

Specific quality changes can be existentially dependent on others. Intuitively this reflects the fact that some quality changes can affect (or be affected by) other changing aspects of that or other endurants. Consider again the case of quality types $Shape$ and $SpatialLocation$. It can be the case that, when an endurant changes its $Shape$ necessarily it changes also its $SpatialLocation$, but not vicecersa. Thus, a change in $Shape$ induces a simultaneous change in the other aspect of the same endurant. On the other hand, one can assume that a change in $Shape$ is necessarily associated with a change in $SpatialLocation$ of a different endurant (occurring in an immediate spatio-temporal proximity) .

On the basis of the temporal relations between quality changes and of the identity relations between their participants, we distinguish three different kinds of *generic existential dependence*[7] between unary predicates (α and β in the following formulas) that individuate sets of quality changes:

(D7) $\mathsf{sQD}(\alpha,\beta) \triangleq \exists x(\alpha(x)) \wedge \forall x(\alpha(x) \rightarrow \exists y(\beta(y) \wedge \mathsf{CN}_T(x,y) \wedge pc(x) = pc(y)))$[8] (*Synchronic Dependence*)

[6]Note that the uniqueness of qualities of a specific type inherent to an endurant (see (Ad44)) guarantees that it is not possible to have different simultaneous basic quality changes of one endurant both relative to the same quality type.

[7]Note that these dependences are more general than the *generic constant dependence* introduced in DOLCE (see (Dd71)) because the last one presupposes the temporal inclusion of the instances of α and β.

[8]It is easy to prove that: $\mathsf{sQD}(PQ_i^C, PQ_i^C)$.

(D8) $\mathsf{bQD}(\alpha, \beta) \triangleq \exists x(\alpha(x)) \wedge \forall x(\alpha(x) \rightarrow \exists y(\beta(y) \wedge y <_T x \wedge pc(x) \neq pc(y) \wedge \phi_1))$
(*Backward Dependence*)

(D9) $\mathsf{fQD}(\alpha, \beta) \triangleq \exists x(\alpha(x)) \wedge \forall x(\alpha(x) \rightarrow \exists y(\beta(y) \wedge x <_T y \wedge pc(x) \neq pc(y) \wedge \phi_2))$
(*Forward Dependence*)

where $pc(x)$ is the participant to a basic quality change, i.e. $pc(x) \triangleq \iota y(\mathsf{UPC_C}(y, x))$[9], and ϕ_j is a formula. In the case above, $Shape^C$ is synchronously dependent on $SpatialLocation^C$ but not viceversa. In addition, $Shape^C$ is backwardly dependent on $SpatialLocation^C$, but $SpatialLocation^C$ is not forwardly dependent on $Shape^C$.

The simple binary dependences are not enough to express complex cases where a quality change depends (backwardly or forwardly) on different simultaneous quality changes of the same endurant. For this cases, we need a notion of *multiple dependence*. Here we introduce only the definition of *multiple backward dependence*, the definitions of the forward and mixed back/forward dependences are similar:

(D10) $\mathsf{bQD}(\alpha|\beta_1, \ldots, \beta_n) \triangleq \exists x(\alpha(x)) \wedge \forall x(\alpha(x) \rightarrow$
$\exists y_1, \ldots, y_n(\beta_1(y_1) \wedge \ldots \wedge \beta_n(y_n) \wedge pc(y_1) = \ldots = pc(y_n) \wedge$
$pc(x) \neq pc(y_1) \wedge \mathsf{CN}_T(y_1, y_2) \wedge \ldots \wedge \mathsf{CN}_T(y_1, y_n) \wedge y_1 <_T x))$
(*Multiple Backward Dependence*)

At this point, we have all the notions necessary to introduce complex dependences corresponding to logical 'or' ($\vee$) of multiple dependences. Intuitively, the logical 'or' represents the fact that the same quality change can 'cause' (or be 'caused' by) other quality changes of a different type. Syntactically, we use a semicolon for the logical 'or', and symbols b (backward) and f (forward) for the kind of multiple dependence. For example, the formula

(D11) $\mathsf{QD}(\alpha|\mathsf{b}\langle\beta_a\rangle; \mathsf{b}\langle\beta_{b_1}, \beta^C_{b_2}\rangle\mathsf{f}\langle\beta^C_{b_3}\rangle) \triangleq \mathsf{bQD}(\alpha, \beta_a) \vee (\mathsf{bQD}(\alpha|\beta_{b_1}, \beta_{b_2}) \wedge \mathsf{fQD}(\alpha, \beta_{b_3}))$

states that α is backwardly dependent on β_a or it is multiple bidirectionally dependent on: β_{b_1}, β_{b_2} (backwardly) and β_{b_3} (forwardly).

In order to specify the 'ontological structure' of qualities and the causal knowledge that we use to define *causation*, we further specialize the previous kind of dependences on the basis of the 'logical/ontological nature' of α, β, and ϕ as discussed in section 3.[10]

Structural dependence. The structural dependence is a synchronic dependence between change types. Structural dependences express very general laws which are not considered causal. They cluster types of changes that are *necessarily* and *synchronously* associated with a given change type involving an object because of its structure (ontological characteristics). In defining *causation*, a *specific* synchronic dependence between quality changes is used:[11]

(D12) $\mathsf{sQD}(x, y) \triangleq \exists \alpha, \beta(\mathsf{sQD}(\alpha, \beta) \wedge \alpha(x) \wedge \beta(y) \wedge \mathsf{CN}_T(x, y))$

Causality dependence. The causality dependences are backward and/or forward dependences between change types where ϕ_j is the proposition *true*. These dependences associate a given type of quality change, occurring to a given object, with quality changes *necessarily* but *non synchronously* occurring to *other* objects (i.e., independently of their ontological characteristics).

---

[9]Note that this definition makes sense only in the case of basic quality changes because, in this case, the existence of an unique participant is guaranteed by (D3).

[10]Generally speaking, we assume the set of dependences in the theory is minimal in the sense that, given $\alpha \neq \alpha'$ and/or $\beta \neq \beta'$, formula $\mathsf{bQD}(\alpha, \beta) \rightarrow \mathsf{bQD}(\alpha', \beta')$ fails. Similarly for $\mathsf{fQD}(\alpha, \beta) \rightarrow \mathsf{fQD}(\alpha', \beta')$.

[11]Note that, once the finite set of synchronic dependences is fixed, this is a first-order formula.

Circumstantial dependence. The circumstantial dependences are backward and/or forward dependences between sets of quality changes in which the formula ϕ takes into account the qualia of the temporal qualities of quality changes and/or the qualia of the physical qualities of their participants. It is possible to distinguish two types of circumstantial dependences.

- *Intrinsic dependence*. Intrinsic dependences are forward and/or backward dependences between subsets of change types. They capture *how* the participant changes its physical qualities, e.g., "temperature increasing backwardly depends on temperature increasing".
- *Relational dependence*. The relational dependences are forward and/or backward dependences between sets of quality changes in which the formula ϕ compares the qualia of the events' qualities and/or the qualia of the participants' qualities. Without discussing their logical forms in detail, here we consider two basic subtypes of relational dependence:
  - dependence involving the qualia of temporal qualities of quality changes (e.g., "the energy of the quality change x must be greater than that of the quality change y");
  - dependence involving the qualia of physical qualities of the participants (e.g., "the final spatial-location of the participant to x must coincide with the initial spatial-location of the participant to y").

4.5 Causation

So far, we have listed several types of dependences. Among these, causality dependences isolate the notion of causality embedded in the overall systems. These dependences, together with the circumstantial ones, allow us to define a notion of (relative) causation, which is a binary relation on events as anticipated in section 3.

First of all, we clarify in which sense the notion of causation we capture is relative. Generally speaking, causation is defined on pairs of events and, as such, its domain is the set of all events in the ontology. In everyday practice instead, causation is a relation applied only to those events that the user considers relevant. That is, one fixes a set of events (perhaps like M_1 or M_2 of section 2) and looks for causation relationships within this set only. Should one define causation relation for any set of events? We know that to establish causation between events a and b, one often has to look for a chain of (pairs of) events that satisfies the causation relationship. This is a crucial issue. Indeed, if one can establish causation between events a and b only through a third event c, the ability to conclude that a, b are in causation relation depends on the presence of c in the given set of events. Thus, the conclusion depends on the given set. One could claim that the given set of events provides all the information one is allowed to use, thus it is from this information only that a conclusion has to be reached. On the other hand, one can sustain that, once other events are known to exist, one has no right to leave them out. This is primarily a modeling decision and one can find real applications, like in legal trials and natural physics, to defend its position. In this paper, we do not take a stand on this issue and leave room for the different approaches. However, in order to show how the theory can deal with this issue, below we apply the causation relationship to sets of events that are closed with respect to the class of *structural* dependences. A generalization to other dependences is easily obtained while it is obvious how to drop this closure restriction.

Let us call *structurally closed* any set of events that satisfies the structural constraints of the system. For the time being, assume that the system includes just one structural constraint which states that for every event a corresponding to a change of shape, there must be a concurrent event a' corresponding to a change of location of the same endurant. If E be an arbitrary set of events, then we can obtain a structurally closed set from E as follows. For each event in E corresponding to a change of shape, add a concurrent event that is a change of location of the same endurant. Call E^* the set thus constructed. Clearly, $E \subseteq E^*$ and E^* is structurally close. We now provide a definition of causation for this kind of sets.

Let E be a structurally closed set. We say that a pair of events (a, b) in E, satisfies the causation relation, i.e., $\mathsf{CS}_E(a, b)$ holds, when at least one of the following occurs:

(i) If a and b are quality changes of type α and β, respectively, i.e. $\alpha(a) \wedge \beta(b)$, then

- a backward or a forward dependence of the form $\mathsf{bQD}(\psi, \phi)$ and $\mathsf{fQD}(\phi, \psi)$, where $\alpha(x) \rightarrow \phi(x) \wedge \beta(x) \rightarrow \psi(x)$, is satisfied, or
- there exist $z_1, \ldots, z_n$ in E of types $\gamma_1, \ldots, \gamma_n$, respectively, such that $\mathsf{sQD}(z_1, a) \wedge$ ... $\wedge\, \mathsf{sQD}(z_n, a)$ and a constraint $\mathsf{bQD}(\psi|\phi, \chi_1, \ldots, \chi_n)$ (or a multiple fQD) is satisfied (where for all x, $\alpha(x) \rightarrow \phi(x), \beta(x) \rightarrow \psi(x), \gamma_1(x) \rightarrow \chi_1(x), \ldots, \gamma_n(x) \rightarrow \chi_n(x)$), or
- one of the remaining constraints expressed following the schema (D11) is satisfied;

(ii) there exist events a' and b' in E such that $\mathsf{sQD}(a, a')$, $\mathsf{sQD}(b, b')$, and (a', b') satisfies condition (i) above ($a = a'$ or $b = b'$ is allowed);

(iii) there exists event c in E such that $\mathsf{CS}_E(a, c)$ and $\mathsf{CS}_E(c, b)$.

From our description, an explicit definition of CS_E can be easily obtained, although the logical formula is a bit complex. In particular, note that our definition is intrinsically recursive. Since in general one cannot bound the number of events in E, it is possible that causation between two events a, b in E can be established only considering a sequence of pairs $(a, e_1), (e_1, e_2), \ldots, (e_n, b)$ whose maximal length depends on E itself.

Note that from this definition (see condition (i)) it follows that $\mathsf{CS}_E(a, b) \rightarrow a <_T b$.

5 What broke the window?

In this section we show how to use our framework to establish causation relationships among events. For this we go back to Example 1 (The Broken Window, see section 2) and provide our analysis of this problem.

The reader has noticed that we have hardly characterized causality and related notions in a normative sense. After all, leaving aside our treatment of causation in the last section, we limited our work to the classification of sets of constraints (structural, causal, circumstantial) and their motivations within the DOLCE ontology. It is the specific choice of constraints that one assumes that isolates the notion of causality and causation one is using and, indeed, we know that there is no unique characterization of these notions people like to use in their everyday work. Thus, in facing a specific example, the initial step is to provide a set of constraints corresponding to (and isolating) the notions of causality and causation that we want to use. Since Example 1 is an example of causation among physical events, we look at classical physics as a source of inspiration and proceed by adopting those physical laws that fit our knowledge about the world. In other terms, in this example we use the laws of physics to relate the qualities available in the system and to add specific constraints to the overall ontology. In addition remember that from (i) of section 4.5, the following is valid: $\mathsf{CS}_E(a, b) \rightarrow a <_T b$. Here, we focus on the study of different possible representations of the situation of the *broken window* informally described in section 2, and analyze how the causality and circumstantial constraints affect the causation relations between specific events inthe universe of discourse.

Our first modeling exercise considers a quite poor ontology, call it $\mathbf{O_1}$, in which just a few quality types are available, namely: spatial location (SL), shape (ShL), and temporal location (TL).The first two, SL and ShL, are physical quality types and TL is a temporal quality type. Using the limited language of $\mathbf{O_1}$, we represent the broken window example introducing three quality changes: a = the canary moves from location l_i^a to l_f^a (a SL change); b = the brick moves from location l_i^b to l_f^b (another SL change); and c = the windows changes shape from s_i^c to s_f^c (a ShL change). In addition we assume $a <_T c$, $b <_T c$, and $a \approx_T b$. This is a very stripped down description of the original problem, however the reader should recognize that, from the point of view of $\mathbf{O_1}$, a, b, c describe 'correctly' the example (in what follows, we consider $W = \{a, b, c\}$). What are the constraints we use in $\mathbf{O_1}$? We want to keep our example simple so let us assume just one constraint in addition to the temporal constraint: since in Newtonian physics a change of shape can be caused by a movement, we state this as a *backward causality dependence* in the theory $\mathbf{O_1}$: $\mathsf{bQD}(ShL, SL)$.

Having specified what $\mathbf{O_1}$ is, we can check which pairs of events fall in the causation relation. The main question is: which event (if any) is the cause of c? First, we must verify whether W is a structurally closed set. Since in $\mathbf{O_1}$ there are no structural constraints, W is closed in this sense and we can apply our definition of causation.Only two pairs of events in W respect the temporal condition. These pairs are: (a, c) and (b, c). Since both a and b are movements and c is a change of shape, then they both satisfy the backward quality dependence and we conclude that both a and b are, independently, cause of the latter event, that is, in $\mathbf{O_1}$ we have $\mathsf{CS}_W(a, c)$ and $\mathsf{CS}_W(b, c)$.

Of course, one may want to add other constraints to the ontology. Following classical physics, one could assume that, in order to exist a causation relationship in the case of movement, the final location of the object in the cause event has to be in contact with the initial location of the object in the effect event. This is a *backward circumstantial* (relational) dependence according to our classification in section 4.4. Let us call $\mathbf{O_2}$ the ontology that assumes the temporal constraints and the latter one. Since $\mathbf{O_2}$ does not include structural constraints, we can apply our causation definition to set W. Because of the temporal constraints, we need to consider pairs (a, c) and (b, c) only. Note that now we must consider the location of the window glass, which is given since it is a quality of the participant to c. Let us call l_i^x and l_f^x the initial and final locations of the participant to the event x and suppose that $l_f^b = l_i^c$ and $l_f^a \neq l_i^c$.The new constraint is satisfied only by (b, c) and thus, in $\mathbf{O_2}$, we conclude that only b causes the braking of the window, i.e., in this theory we have $\mathsf{CS}_W(b, c)$ but not $\mathsf{CS}_W(a, c)$.

Instead of the new constraint on locations, used to obtain $\mathbf{O_2}$, one could consider a new quality type, for example a quality type providing the energy of an event. From classical physics, we know that an event can cause another only if the first can deploy at least as much energy as needed by the latter.[12] Let us add this constraint to $\mathbf{O_1}$ and call the resulting ontology $\mathbf{O_3}$. As before, we need to check pairs (a, c) and (b, c) only. This time we must include in our analysis the energy of each event, energy that here we indicate with ϵ^a, ϵ^b, and ϵ^c. To establish which event(s) is cause of c in $\mathbf{O_3}$, if any, we need to check if $\epsilon^a \geq \epsilon^c$, for (a, c), and if $\epsilon^b \geq \epsilon^c$, for (b, c). The result will establish which of $\mathsf{CS}_W(a, c)$ and $\mathsf{CS}_W(b, c)$, if any, is true in this ontology.

Another interesting case is obtained by changing the universe of discourse. Let us substitute event a by $a^* =$ the canary changes shape. Also, this time we include a *structural* constraint stating that in order to have a change in shape a movement must occur. What is the causation relation in the resulting ontology? From our discussion in section 4.5, one cannot consider the set of events $W^* = \{a^*, b, c\}$ since it does not satisfy the structurally closure constraint. One must look at the richer set $W^+ = \{a^*, b, c, d, e\}$ that includes the new events $d =$ the canary moves and $e =$ the window glass moves. The reader should note that W^+ is uniquely identified by W^* and the structural constraints we have adopted. At this point, ac-

[12] The case of concurrent events that are co-cause is captured by considering their mereological sum.

cording to the remaining constraints in the ontology here considered, one can apply the given definition of causation (now with respect to the new set W^+) and argue as in the previous cases to conclude $\mathsf{CS}_{W^+}(d,c)$ and, from this, $\mathsf{CS}_{W^+}(a^*,c)$ or, alternatively, not $\mathsf{CS}_{W^+}(d,c)$ and not $\mathsf{CS}_{W^+}(a^*,c)$.

At this point, it should be clear how different ontologies isolate different notions of causality and causation in our framework.

6 Conclusions

We have discussed the definitions of causality and causation relations and their formal and applicative drawbacks described in the literature. Recognizing the validity of such criticisms, we have begun to study how constraints of different nature (structural, causality, and circumstantial) intervene in shaping causal relations. We proposed to look at these constraints as forms of dependences among event types that capture general laws (at different levels of abstractions) or present orthogonal aspects of these relationships. This study allowed us to propose different definitions of causality and causation which depend on several parameters like the adopted ontology (in particular the qualities it includes), the set of events taken into consideration as well as special dependences among events and their types.

We remark that our proposal is admittedly incomplete. For instance, it does not make justice of the intuitive notion of causation in the case of static events: the pen on the table does not fall "because" of the presence of the table. Many argue that this case involves a truly causal relationship. This is not clear to us. Anyway, the approach presented in this paper is not applicable to such cases. Also, our work does not explain how to extend the proposed framework to complex events. Although one can consider complicated sets of events by applying the theory to their mereological sum, the effectiveness and clearness of the resulting causality and causation relations need to be investigated further.

References

[1] C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. Wonderweb deliverable d18 - final report. Technical report, National Research Council - Institute of Cognitive Science and Technology, 2003.

[2] J. Lehmann. *Causation in Artificial Intelligence and Law - A modelling approach*. PhD thesis, University of Amsterdam - Faculty of Law - Department of Computer Science and Law, 2003.

[3] J. Lehmann, J. Breuker, and B. Brouwer. Causation in ai&law (to appear). *AI and Law*, 2004.

[4] C.J. Ducasse. On the nature and observability of the causal relation. *Journal of Philosophy 23 57-68*, 1926.

[5] J.S. Mill. *A system of logic ratiocinative and inductive*. London, 1886.

[6] J.L. Mackie. *The cement of the Universe - A study in causation*. Oxford University Press, 1974.

[7] J. Pearl. *Causality*. Cambridge University Press, 2000.

[8] D. Lewis. Causation. *Journal of Philosophy 70 556-567*, 1973.

[9] P. Dowe. Causality and conserved quantities: A reply to salmon. *Philosophy of Science 62, 321-333*, 1995.

[10] D. Fair. Causation and the flow of energy. *Erkenntis 14 219-250*, 1979.

[11] B. Russell. *Human Knowledge*. Simon and Schuster, 1948.

[12] W. Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press, 1984.

[13] J. Allen and P. J. Hayes. A common-sense theory of time. In *International Joint Conference on Artificial Intelligence, IJCA'85*, 1985.