Causation in AI&Law

Jos Lehmann (joslehmann@ip.rm.cnr.it)
Laboratory for Applied Ontology, Institute of Cognitive Science and Technology, Italian National Research Council, Rome, Italy

Joost Breuker (breuker@lri.jur.uva.nl)
Leibniz Center For Law, Faculty of Law, University of Amsterdam, Amsterdam, The Netherlands

Bob Brouwer (p.w.brouwer@uva.nl)
Department of Jurisprudence, Faculty of Law, University of Amsterdam, Amsterdam, The Netherlands

Abstract. Reasoning about causation in fact is an essential element of attributing legal responsibility. Therefore, the automation of the attribution of legal responsibility requires a modelling effort aimed at the following: a thorough understanding of the relation between the legal concepts of responsibility and of causation in fact; a thorough understanding of the relation between causation in fact and the common sense concept of causation; and, finally, the specification of an ontology of the concepts that are minimally required for (automatic) common sense reasoning about causation. This article offers a worked out example of the indicated analysis. Such example consists of: a definition of the legal concept of responsibility (in terms of liability and accountability); a definition of the legal concept of causation in fact (in terms of the initiation of physical processes by an agent and of the provision of reasons and/or opportunities to other agents); CausatiOnt, an AI-like ontology of the common sense (causal) concepts that are minimally needed for reasoning about the legal concept of causation in fact (in particular, the concepts of category, dimension, object, agent, process, event and act).

Keywords: Legal responsibility, causation in fact, common sense, ontology

1. Introduction

This article presents the most relevant results of an AI&Law-like research project (Lehmann, 2003), which was envisioned as part of the Functional Ontology of Law (FOLaw, in the following) (Valente, 1995), between 1991 and 1997 at the Department of Computer Science and Law of the University of Amsterdam. Between 1998 and 2003 the project was realized at the same department, recently renamed Leibniz Centre for Law. The central topic of the reported investigation is the representation of causation for automatic legal reasoning. By expanding FOLaw’s with modules for Causal and Responsibility knowledge, the main purpose of this study is the definition of an analytical model of the concepts (and the conceptual relations) used in legal reasoning when
assessing causation in a case description. Such a model may be used as a basis for automatic legal reasoning. In this sense the link between the research presented in this article and AI research is Ontology, the most philosophical subfield of Artificial Intelligence, which has steadily been growing for the last fifteen years.

The paper is structured as follows. In the rest of this introduction we provide a first intuitive treatment of causal issues in legal settings: we provide arguments in favor of considering causation in fact as a relevant problem both for Legal Theory and for Artificial Intelligence and Law. Section 2 concentrates on the legal (theoretical) concept of responsibility and on the notion of ground for the attribution of legal responsibility. One of such grounds is causation in fact. Section 3 concentrates on the legal theoretical notion of causation in fact. Three main legal theoretical approaches to causation in fact are presented: causal maximalism, causal minimalism and Hart and Honoré’s approach. This last approach is chosen as our legal theoretical point of reference on matters of causation in fact. Section 4 presents a preliminary analysis of the ontological elements that are implicit in Hart and Honoré’s proposal. Their definition refers to four ontologically distinct causal relations: physical, agent, interpersonal and negative causation. Section 5 concentrates on the cases of physical and agent causation and provides an overview of the main problems that should be considered when defining the relations of physical and agent causation. Section 6 presents CausatiOnt\(^1\), an ontology developed as a basis for modeling physical and agent causation. For reasons of conciseness, we do not present all the ontology nor a fully worked out application of the ontology to the canonical example used in this paper. Our overall concern in this paper is to show how, through philosophical analysis, a well founded legal theoretical treatment of the notions of responsibility and of causation in fact may be turned into (the fragments of) a rich AI-like ontology of causal concepts. The full specification and application of CausatiOnt would require a paper on its own and, to achieve it here, we should cut short on the legal theoretical foundations of the ontology. But this would undermine exactly the interdisciplinary approach that characterizes this paper. Therefore, for a fully specified treatment of CausatiOnt the interested reader is referred to chapter 3 of (Lehmann, 2003). Finally, Section 7 draws some conclusions.

\(^1\) From CAUSATIon ONTology, indicated as \textit{CAUSATI}ONT in headings and captions throughout this article
1.1. LEGAL KNOWLEDGE ABOUT CAUSATION

Before getting started with the presentation of the contents of our model - i.e. before addressing the what-questions - we would like to deal with a few methodological issues - i.e. a few why-questions. Such cautious move is primarily aimed at clarifying why (we believe that) the automation of reasoning about (legal) responsibility cannot be achieved without the automation of (legal) causal reasoning (or of some of its fragments). The next two subsubsections are thus primarily dedicated to the followers of those legal-theoretical schools that maintain that there is no functional or analytical dependence between legal responsibility attribution and legal causal reasoning. Subordinately, though, the convinced causalist may too profit of the material presented in the following, by getting an idea of the main arguments in favour and against causal views on legal responsibility.

1.1.1. Causation in fact as a problem for (automatic) legal reasoning

A significant part of the AI and Law community has strong doubts on whether causation is a problem at all for legal reasoning. These doubts may be traced back to the classical distinction between causation in fact, on the one hand, and legal causation, on the other hand. Causation in fact is the problem of understanding what actually happened (i.e. what caused what) in a case. Such factual interpretation is something legal experts usually take for granted and mostly see as unproblematically achieved by common sense. On the contrary, they see legal causation as problematic and, therefore, interesting. Legal causation is the set of criteria that should be applied either when a clear common sense factual interpretation of the case is missing or when, despite having a clear causal interpretation of the case, legal policy considerations should be applied (e.g. foreseeability) and this results in adopting a causal interpretation that is different from the factual causal one. Typical examples of cases where a legal causal interpretation must be used, because a factual interpretation is missing, are so called cases of overdetermination. For instance, two shooters hit the same one person at the same time and each shot is such that it may have caused her death. Who caused her death? From a causal factual point of view, further investigation is needed in order to establish which of the shooters actually caused the person’s death. And even such an investigation may have no more means than common sense has for reconstructing the actual chain of events. Therefore, from a legal causal perspective it is solomonicly\(^2\) assumed that both shooters caused the death of the person. On the other hand, typical examples

\(^2\) With justice, like Solomon would have done.
of cases where a legal causal interpretation substitutes a clear factual interpretation are cases where foreseeability or risk play a role. For instance, a truck-driver transporting toxic substances has an accident and the toxic substances are spilled on the road. Subsequently, the toxic substances drain through the soil into a groundwater reservoir, thereby polluting the water and creating great damage to a nearby village’s water supply. Despite the clear factual connection between the truck-driver’s conduct and the polluting of the water, from a legal causal perspective one may argue that the truck-driver did not cause the polluting, because he had no means of foreseeing that there was a groundwater reservoir right at the place where the accident took place. In other words, in the legal analysis of such difficult cases as those illustrated above, questions of factual causation are simply left unanswered or their answers are set aside. This has resulted in a skepticism of a more general nature concerning the very necessity of treating causation in fact as a legal theoretical problem. “Don’t worry - say the skeptical legal theoreticians (Green, 1930) - be happy! Because causation in fact is only a minor detail. Legal reasoning has all it needs in order to determine (and attribute) legal responsibility without going into the details of what caused what.”.

An intuitive refutation of this position, may be summarized in the phrase dubitant ergo est - they doubt, therefore it exists. Such phrase is the conclusion of the following counterargument. The problem of consistently and completely defining the relation of causation between two events or states of affairs is one of the hardest among the scientific and philosophical questions of all times. Every theory of the world (e.g., Physics, Chemistry or Philosophy) and any organized corpus of knowledge about the world (e.g., the Law, Medicine or History) suffers from internal paradoxes which are provably due to some lack of understanding of causal relations\(^3\). The causal plague is so diffused and general that each scientific field has developed its own particular rules of the thumb for dealing with causal issues. As a matter of fact, specific causal rules have so far proven to be much more reliable than general inferential schemes. The latter should in principle guarantee more uniformity in distinguishing causal from non causal relations; but the former actually provide a tighter correspondence between general causal intuitions and each field’s peculiar way of describing basic facts, of assessing their truth and of handling exceptions. This sort of “causal scientific secessionism” has generated a lot of technical views on causation which are hardly comparable to each other. Nonetheless, as the

\(^3\) In this article, for a philosophical causal paradox see Example 1: counterfactuals generate pretty absurd interpretations; for a legal case, which is particularly tough even though not paradoxical, see Example 2.
word cause and its synonyms keep being used for communication within and across different fields, secessionism has also produced a general desire for unity in the form of “federalism”. Every sufficiently theoretical study of causation almost invariably makes two statements:

1. The word “cause” refers to a cluster of variegated and sometimes even conflicting concepts, rather than just one clear-cut notion.

2. The unity of such diverse conceptual family is guaranteed by common sense, i.e. by something that everybody has (i.e., common) but that nobody can rationally describe (i.e., sense).

The law makes no exception, neither for its sometimes provincial tendency to secessionism nor for its desire to fully take part to the causal federation. Going through the legal-theoretical debate on the problem of causation makes very clear how deep the dichotomy is between different groups of legal theoreticians. Some, the secessionists, sincerely think that legal practice and reasoning has hardly any problems with causation in fact. In their eyes, in most cases the law has to do with attributing legal responsibility; in other words, legal reasoning deals only with so-called questions of policy. On the other hand, the federalists try to convince their fellow scholars of the following: if it is true that legal practice has felt the need to let juries or judges solve difficult questions of fact arising in legal cases, then it is also true that legal reasoning has problems with causation in fact as much as any other type of reasoning has. Therefore, the pragmatic solution (e.g., letting juries decide questions of fact) should not be considered as a miraculous way out of the problems of causal reasoning. Juries do not separate the courtroom from the causal federation. On the contrary, they work as the channel of communication between the legal apparatus and social common sense, in any given moment of time. The function of juries is twofold: they represent in the courtroom general social knowledge (at least in the form doubts, if not as certainties) about questions of fact and they provide a reasonable solution for the specific case they are called to judge. In short, juries at least doubt; therefore, the problem, that they are called to doubt about, exists. *Dubitans ergo est.*

A second argument, in favour of considering the relations between the factual-causal interpretation of a case and the attribution of responsibility as very tight relations, is an argument by example, from (Pearl, 2000).

**EXAMPLE 1 (The Desert Traveler).** A desert traveler T has two enemies. Enemy 1 poisons T’s canteen and Enemy 2, unaware of Enemy 1’s action, shoots and empties the canteen. A week later, T is found dead and the two enemies confess to action and intention.
If a jury were asked to attribute the legal responsibility for T’s death, it would probably have to consider the following additional information, which is left implicit in Example 1:

1. T never drank from the canteen.
2. T was found dead by dehydration.

Based on such information, the jury would very probably come to an unanimous decision and indicate Enemy 2 as the responsible person for T’s death. If asked why, the jury may answer: because Enemy 2 caused T’s death. If asked in what sense Enemy 2 caused what he caused, the jury would probably say that Enemy 2’s action is a *counterfactual condition* of T’s death, which makes it a cause. In other words, had Enemy 2 not shot the canteen, T would still be among us. But this is not true - it should be replied. Had Enemy 2 not shot the poisoned canteen, T would have drunk from it and he would not be among us anyway. Therefore, Enemy 2’s action is not a counterfactual condition of T’s death. Is it still its cause? - the jury should be asked. Again its answer would probably be unanimous and indicate Enemy 2’s action as the cause of T’s death in the sense that he is the most *proximate cause* of T’s death. If asked to give a definition of such proximity, the jurors would probably give a temporal definition of such proximity: Enemy 2’s action is the latest cause of T’s death. But, then again, it could be replied that from a strictly physical point of view the heat of the Sun was definitely a temporally more proximate cause than Enemy 2’s action.

This “cat and mouse game” with the jury could go on for a long time. This is due to the fact that Example 1 is no real-life case. It is just a tricky and underspecified combination of circumstances devised by some smart philosopher on some lazy day, with the explicit purpose of fooling imaginary juries. The example, though, does show at least one thing: a “short circuit” in our causal understanding of a series of events has major consequences on our capacity to attribute (legal) responsibility.

Finally, (Aqvist and Mullock, 1989), (Prakken and Renooij, 2001) and (Baldwin and Neufeld, 2004) offer more arguments in favor of recognizing causation in fact as a major question for AI and Law. All these studies show how automatic legal reasoning would greatly profit from a causal upgrade. We will make no further use of these references because their approaches are somewhat orthogonal to our own and treat the problem of causation in fact as a problem of inference. They propose the adoption of complex game-theoretical, logical and, respectively, probabilistic machinery for tackling the problem, whereas we take a
chiefly ontological stance.
In conclusion, it seems convenient, if not essential, to develop AI and Law-like theories of causal relations. This would make automatic legal reasoning both more efficient (i.e. speeding up the software performance) and more effective (i.e. refining the software understanding of the world and enhancing its reasoning power about cases and/or legislations).

1.1.2. Evidence and the burden of proof
The second methodological issue we treat here is why our treatment of causation in fact makes abstraction of problems of evidence and of the burden of proof. As partly explained in (Hart and Honore, 1985) (pp. 406-428), those two problems influence the role of legal causal reasoning only by regulating the acquisition of information by the court. This means that in principle, as well as in most real cases, evidence and the burden of proof affect legal causal reasoning solely from a procedural point of view. They do not inherently change the kind of information that is considered by a court when assessing causal relations in a case. At most, problems of evidence or of the burden of proof may affect the amount or the order of the information provided to a court for making a decision. Despite the fact that incomplete or poorly ordered information may lead a court to make unjust decisions, this does not affect the type of information that a court requires in order to make a decision, be it just or unjust.

In other words, throughout our treatment of causal issues in legal settings we will rely on a complex assumption by which it is assumed that whoever between the parties has the burden of proof:

1. Sufficient evidence is provided that each considered fact or event actually took place. We therefore assume, for instance, that in Example 1 there is sufficient evidence that it was Enemy 2 who pulled the trigger.

2. Sufficient evidence is provided that no relevant facts have been omitted. We assume, for instance, that Enemy 2 was not forced by someone to fire.

1.1.3. Scientific problems
Provided that the problems of legal reasoning about causation in fact are relevant problems, what kind of scientific problems are they? What kind of questions do they raise? Rather difficult questions, must be said. As a conclusion to this introduction we provide a first schematic overview of the main groups of what-questions that are treated in sections 2 through 6.
Group 1: Legal responsibility  The question must be answered what is meant by legal responsibility. What is the role of this concept in legal reasoning? And what is its relation with other similar legal concepts such as liability and accountability?

Group 2: Legal responsibility and causation in fact  The relation between causation in fact and legal responsibility must be clarified. Is causation in fact part of the definition of legal responsibility? Or is it a separate concept which somehow interacts with the concept of responsibility?

Group 3: Causation in fact  Questions arise concerning the relations between the legal and the common-sensical notion of causation in fact. Is the former a specialization, or an expansion or a “recombination” of the defining elements of the latter?

Group 4: Causation  After answering the questions on the legal notion of causation in fact, the realm of Legal Theory must be abandoned for the open territory of Philosophy. A definition of common-sense causation is needed. And this in turn raises a lot of philosophical questions. Can we distinguish between different types of causation (e.g., physical vs agent causation)? How? Is causation a logical (and deterministic) relation? Or is it a probabilistic (and indeterministic one)? Or is it yet another type of relation? What does causation relate: what are its relata? Is causation (and its types) the only causal relation?

Group 5: Representation of causation  Representing causation in a machine-friendly way raises yet another group of questions. In particular the problem arises of how to represent causation in a parsimonious way, i.e. in a way that is computationally not too demanding.

Section 2 treats questions pertaining to Group 1 and 2; section 3 treats questions pertaining to Group 3; sections 4, 5, 6 treat questions pertaining to Group 4 and 5.

2. Legal Theory on responsibility

The present section elaborates on the legal (theoretical) concept of responsibility and on the notion of ground for the attribution of legal responsibility.
2.1. **The air rifle example**

The theoretical literature about causal relations is riddled with an exceptional variety of disparate examples. We too adopt a parametric example, which helps us in the illustration of various theories that are relevant to our purposes. We use here a case taken from the legal theoretical literature. Due to the many complications that characterize it, this example illustrates both the pervasiveness of causal relations in our concrete experience and the elusiveness of such relations when trying to conceptualize them.

**EXAMPLE 2** (The air rifle). In breach of a statute forbidding the sale to an infant under the age of 16 of dangerous weapons, the defendant sold an air rifle and ammunition to a boy of 13. The boy’s mother told the boy to return the weapon to the defendant and get a refund: on the defendant’s refusal to take the rifle back, the boy’s mother took it from the boy and hid it. Six months later the boy found it and allowed a playmate to use it, who shot and accidentally wounded the plaintiff, destroying the sight of one eye.

(Henningsen v. Markovitz (1928) 132 Misc. 547, 230 NYS 313)

The dramatic development of the events described in Example 2 raises a general question of responsibility. This is due to the fact that the final event described in the example is a harmful one. It is mostly - even though not exclusively - about undesirable events that people seriously ask themselves: "Who made this happen? Who is to blame?". Most people would agree that there is at least a question of moral responsibility connected to the described chain of events. The attribution of moral responsibility, though, may greatly vary from individual to individual. Someone might see the mother as morally responsible of the plaintiff’s lost eye; someone else might have reasons for pointing his finger against the seller (the defendant); someone else might consider morally responsible the manufacturer of the air rifle or the engineers who designed it. There even might be someone who blames the victim himself, seeing in his blindness a sort of payoff decided by a metaphysical entity (e.g. God) based on mysterious (but necessary) criteria of (divine) justice.

Now, if the attribution of responsibility (for the plaintiff’s lost eye) is not simply of a moral nature, but of legal one, then the great variation of judgement indicated above may turn into a very serious practical problem. As shown in (Lehmann, 2003) (p. 20), the legal practice normally relies on an assumption of consistency and of uniformity of application (of any given norm). Therefore the attribution of legal responsibility requires detailed (and preventive) analyses concerning the explicit criteria that should be used by a judge when blaming someone for the harm inflicted on someone else. It is precisely such longing for
consistency that raises tough problems concerning the criteria of legal responsibility attribution, problems that are further illustrated in the next subsection. A closer and more systematic look is given at various aspects of Example 2 that are relevant to the attribution of legal responsibility. Legal Theory provides different and alternative types of criteria for legal responsibility attribution. Such a variety shows that moving from the widely varying views on moral responsibility attribution to a more structured and consistent legal view requires a very complex analytical process.

2.2. LEGAL RESPONSIBILITY

Most legal-theoretical debates about the notion of responsibility revolve around the problem of how much space common-sense has in the determination and the attribution of legal responsibility. A case presents two main types of elements, which may play a role in attributing responsibility.

1. On the one hand, there are factual elements, which contain information for establishing the chain of causation and which, therefore, make it possible to attribute the responsibility of the harm to the person(s) who caused it.

2. On the other hand, there are the legal elements of the case, which contain information for identifying the person(s) who may be held responsible for the harm, based on so called considerations of legal policy.

Now, as observed in (Hart and Honore, 1985) (p. xlviii), most legal theoretical treatments of legal responsibility are not neutral with respect to the roles that either causation or legal policy play in the attribution of responsibility. This situation makes the overall debate on responsibility very difficult, due to the lack of a common ground for comparing different approaches.

In order to overcome such lack of a neutral ground of confrontation, Hart and Honoré propose to see legal responsibility as described by three main elements. These form together a general framework for treating issues of responsibility, without requiring any strong commitment about the role of causation or of legal policy. The three elements indicated by Hart and Honoré are: a definition of the legal concept of responsibility, the legal grounds for the attribution of legal responsibility and, finally, the types of cases that arise from combining different grounds. The main reason for adopting Hart and Honoré’s three-tiered approach as our legal theoretical reference on responsibility, is to guar-
antee to our treatment a high level of generality. In the following two
subsubsections, we illustrate this in more detail.

2.2.1. The legal concept of responsibility
Hart and Honoré (Hart and Honore, 1985) (p. xliii) reduce the notion
of legal responsibility to the legal status of someone who is subject to
a legal punishment or a sanction.

DEFINITION 1 (Legal responsibility). Legal responsibility is the lia-
bility of a person to be punished, forced to compensate, or otherwise
subjected to a sanction by the law.

The definition above does not give any indications regarding what a
person must have done in order to be held legally responsible. Hart
and Honoré limit the concept of legal responsibility to the purely le-
gal aspects of this notion and they reduce legal responsibility to the
legal status that a person acquires (i.e. the liability to certain dis-
agreeable consequences). Analytically speaking, this definition might
sound ill-founded because it is given in terms of (legal) consequences.
One would rather expect a proper definition to spell out the legal, non
legal, necessary and/or sufficient conditions that a person must fulfill
in order to be considered responsible (e.g. if you commit a murder,
then you are legally responsible for the death of the victim). Hart and
Honoré’s approach is not that odd, though, when seen in the context
of legal-theoretical debates. This is true for three reasons.

1. First of all, by binding the notion of legal responsibility to the
notion of liability Hart and Honoré mark a clear distinction be-
tween legal responsibility and any other form of responsibility, in
particular moral responsibility. This is due to the very meaning
of liability. The liability of a person may be seen as the person’s
relation with a (judicial) authority. Such authority has the power
to make decisions that directly affect the person and her future.

2. Secondly, defining legal responsibility in terms of liability entails
the presence of a second legal requirement: the accountability of the
person. In order to be liable, a person must be accountable, which
depends on whether the person satisfies the criteria of accountabil-
ity fixed in the law. These criteria usually refer to the physical and
psychological capacity of a natural person to have control on her
actions and/or to stand trial. Typical examples of such criteria are
age or mental sanity.

3. Thirdly and lastly, limiting the definition of the legal concept of re-
ponsibility to liability has another legal-theoretical merit. It shields
the definition of legal responsibility from the (legal-theoretical) debates about the existence of necessary and/or sufficient conditions for legal responsibility. Hart and Honoré propose instead to see such so-called conditions as grounds, which are used in attributing legal responsibility, rather than as defining conditions of legal responsibility itself.

Having clarified the legal nature of responsibility, we illustrate the grounds that are usually used by courts in attributing legal responsibility and the typology of cases generated by combining such grounds.

2.2.2. Grounds for responsibility attribution and types of cases

According to Hart and Honoré: “There are many grounds on which responsibility may be imposed, and others may be invented in the future.” (p.xliv). In other words, the grounds for responsibility attribution do not have the status of logical or strictly rational conditions. They rather are widely accepted requirements, which generally grow out of tradition and that are progressively codified by legislators in the Law. The difference between a logical condition and a requirement may be understood in terms of the (Kantian) difference between analytic statements (logical conditions) and synthetic statements (requirement). As explained above, such requirements should therefore be left outside the definition of legal responsibility. They may instead be spelled out in a separate definition, like the following one.

**DEFINITION 2 (Grounds for legal responsibility attribution).** *Grounds for the attribution of legal responsibility to a person for a given harm are:*

1. **The conduct of the person.**

2. **The causal connection between the conduct of the person and the given harm.**

3. **The fault legally implied by the conduct of the person.**

Definition 2 revolves around the notion of conduct. This is not further defined by Hart and Honoré, but it can be quite safely taken as indicating the intentional or unintentional behavior of a person throughout the events under analysis. Furthermore, the definition considers both factual and legal elements as grounds for the attribution of responsibility. This is in accordance with Hart and Honoré’s intention of giving equal consideration to causation in fact and legal policy in their framework definition of legal responsibility. In Example 2 an instance of conduct is the combination of the seller’s sale of the rifle and his later refusal
to accept it back; an instance of fault is the seller’s breach of the mentioned statute; finally, an instance of a causal connection is the relation between the shooting and the blinding of the victim.

By combining the grounds mentioned above, a wide variety of actual cases may be described. There are five main types of cases:

**Type 1: Conduct, Causation in fact, Fault** In this case type (which exactly corresponds to Example 2) the court allocates responsibility based on the causal chain that started from the seller’s faulty conduct and led to the harm. In other words, the court has to decide whether the unlawful behavior is causally related to the harm.

**Type 2: Conduct, Causation in fact** Suppose that in Example 2 the seller’s conduct, despite resulting in a violation, is not faulty, because it is proven that the boy who bought the rifle showed a faked identity card and the shop keeper had no reasons to assume that this card was faked. The seller could still be responsible for the plaintiff’s blinding based on strict liability and on the causal connection between his conduct and the harm. In other words, in this case the court may allocate responsibility to the seller, based on the causal connection between his conduct and the harm and on his strict liability as a seller of dangerous implements. Furthermore, the strict liability of the seller could be used as an argument against him even if his non faulty conduct did not result in a violation. The seller could be considered responsible for the harm caused to the plaintiff, simply based on his causal involvement in the case.

**Type 3: Conduct, Fault** Suppose that in Example 2, no harm takes place (i.e., there is no causation involved at all). In this case a court could only allocate responsibility to the seller for his faulty conduct, i.e. his breach of the statute. This, by the way, actually happened: before the case presented in Example 2, the seller was actually convicted by a criminal court for his breach of statute.

**Type 4: Conduct** Imagine a combination of types 2 and 3, where no harm takes place and the seller is not to blame because it is proven that the boy who bought the rifle did it by showing a faked identity card. The court could then allocate responsibility to the seller, based on his causal connection with the harm.

---

Hart and Honoré indicate only types 1 to 5 as the most important types of cases, based on empirical considerations rather than on combinatorics. According to the authors, actual cases normally fall under one of these types. There is, though, a more profound reason why combinations such as ‘no conduct-causation in fact-fault’, ‘no conduct-causation in fact-no fault’, ‘no conduct-no causation in fact-fault’ find no consideration. It is legitimate to think that, from a legal theoretical stance, both causation in fact and fault imply (or depend on) conduct, i.e. agency.
card and the shop keeper had no reasons to assume that this card was faked. In this case a court could still allocate responsibility to the seller for his conduct, simply based on his strict liability as a seller of dangerous implements.

**Type 5: No Conduct, No Causation in fact, No Fault** This last type of case cannot be exemplified by directly considering the seller, because he has a role in the case. Clear examples of persons that could be found legally responsible for the harm, despite having no role in the case (vicarious liability), are the parents of the boy who shot (i.e. the playmate), given that there was no omission/negligence on their part.

Now, according to Hart and Honoré’s analysis two types of cases require the assessment of causal relations as a ground for attributing legal responsibility (the first and the second type above). Therefore, if cases were evenly distributed over their types, we would have to conclude that at least two cases out of five require a thorough and coherent analysis of causal relations when a court decides on the related questions of responsibility. We believe, though, that the amount of cases, for which causal reasoning is required, goes well beyond two fifths of the total cases. This is true for the following reasons:

1. *It is very improbable that the distribution of cases over their types is even.* Most cases discussed in court deal with actual undesirable events, which were brought about by someone’s (possibly faulty) conduct.

2. Even when considering cases of the third and fourth types, where no actual harm has happened, courts may have to make (quite extensive) use of hypothetical causal reasoning. The (ontological) basis of such hypothetical causal reasoning are exactly the same as the basis of reasoning about actual causes.

3. For what concerns the fifth type of cases indicated by Hart and Honoré, it is quite evident that the responsibility of a parent to pay for the damages caused by the children, can only be attributed once it is clearly established that the children actually caused the damages. Which again imposes to the court to assess the causal grounds on which the decision is taken.

The situation described above imposes the formulation of clear criteria for handling problems of causation in fact when dealing with questions of responsibility. Legal theory has indeed dedicated a lot of attention to define such criteria. Three main (families of) approaches
may be individuated in this stream of legal theoretical research: causal maximalism, causal minimalism and Hart and Honoré’s approach. The following section illustrates each of these approaches in some detail.

3. Legal Theory on causation in fact

This section concentrates on the legal theoretical notion of causation in fact. Three main legal theoretical approaches to causation in fact are presented: causal maximalism, causal minimalism and Hart and Honoré’s approach. The last approach is then chosen as our legal theoretical point of reference on matters of causation in fact. We invite the reader to mentally try out the presented approaches on Example 2 every time this is not already done by us (for reasons of conciseness).

3.1. Causal maximalism

The most traditional view on problems of legal causation allows to trace legal responsibility through causal relations solely: the person who caused the harm is responsible for it. The causal maximalist recognizes only clause 2 of Definition 2 as valid ground of responsibility attribution. This approach does not allow the use of any other consideration, besides purely causal ones. Therefore, Hart and Honoré call this view causal maximalism, because it simply equates legal responsibility with causation in fact. Causal maximalism is much stronger than the position we expressed at the end of section 2. There we stressed that the causal relation is the central (or paradigmatic) criterion for responsibility attribution. Causal maximalism goes well beyond this and sees the causal relation as the only admissible criterion of legal responsibility attribution. This, of course, leaves the causal maximalist with the open problem of establishing clear criteria for individuating causal relations in legal settings. A few examples of maximalist criteria are:

Causal proximity The most general criterion provided by causal maximalists is a combination of common sense and of the principle of so called causal proximity to the harm. In other words, the agent that by common sense may be considered as the most proximate cause of the harm is the cause of the harm and should, therefore, be held responsible for it.

Beale’s criteria These rules reformulate the notion of proximity in terms of mechanical forces, by providing a sort of Newtonian universal view on causation in fact.
Epstein’s criteria Inspired, on the one hand, by Beale’s notion of force and, on the other hand, by Hart and Honore’s analysis, which was firstly presented in the 1959 edition of (Hart and Honore, 1985), Epstein’s criteria reformulate the notion of proximity in terms of ‘paradigms’: the use of force; fright; the exercise of compulsion; and the creation of a dangerous situation (Hart and Honore, 1985) (p.lxxiv).

The most serious criticism against causal maximalism and its extensions concerns their transparency. As Hart and Honoré put it: “Rules for determining proximate cause very often contain explicit or implicit reference to principles of legal policy.” (Hart and Honore, 1985) (p.97). It must be said that the two English scholars do not fully agree with such harsh criticism. They nonetheless present it because it bears some truth, which justifies the rising of a second legal theoretical movement, known as causal minimalism, that opposed causal maximalism and finally took its place in the harts and minds of many legal theoreticians and practitioners.

3.2. Causal minimalism

In the 30s a new wave of legal theoreticians, lead by Leon Green (Green, 1930), opposed the very possibility of distinguishing between questions of fact and questions of policy. They claimed that almost every issue in a legal case may be reduced to a question of legal policy. The causal minimalist recognizes only clause 3 of Definition 2 as a valid ground for responsibility attribution. These should therefore receive most of the attention, because once the questions of policy of the case are solved, the remaining questions of fact may be tackled by (standard forms of) counterfactual reasoning.

This approach has - so to speak - a nice refreshing flavor if compared to causal maximalism. It cuts loose from all the complications connected to the maximalist’s search for what is causally proximate to (or remote from) the harm. Apparently, the combination of a thorough legal analysis and a counterfactual test is much more practical than (pure) causal analysis. This is due to two intertwined reasons. On the one hand, such combination allows to clarify by purely legal means the relative positions of the persons involved in the case. This very fact, on the other hand, allows to separately assess the causal influence of each person’s conduct by applying purely logical means of analysis (i.e. counterfactuals).

The tests proposed by the minimalists are:

Sine qua non test The first and most traditional counterfactual test formulated for application to legal analysis is the sine qua non
test. The most distinctive point of this test is that it checks for necessary conditions. In other words, according to this logical test all conditions that are recognized as necessary for the harm should be considered as causes of the harm. On its turn, the notion of necessity that supports this view is usually reduced to Mill’s philosophical conception of cause. According to Mill the cause of an event (or, rather, of an event of a certain type) is the sum total of all the conditions which are together sufficient to produce it, in the sense of being invariably and unconditionally followed by it (Mill, 1886). Each of these jointly sufficient conditions is necessary to the effect.

**But-for test** The so called but-for test helps focusing the inquirer on the sufficient rather than on the necessary elements of the causal relation. The question that must be asked is, again, a counterfactual one, ranging on the conducts of the persons involved in the case. This test, though, is phrased in a slightly different way than *sine qua non* counterfactuals: “Would the harm have taken place but for the conduct of this person?” A negative answer is equivalent to the assessment of a causal relation.

**Probability tests** There is a group of tests which detects causation in fact by testing the increased probability of the harm given the conduct of a certain person. These probabilistic approaches retain a counterfactual flavor, even though this is not always patent. In a purely probabilistic approach, an event causes another event if the occurrence of the first event is followed with high probability by the occurrence of the second event. This is very similar to the counterfactual statement that, if the first event had not occurred, the second would not have either. This similarity with counterfactual approaches makes probabilistic ones an interesting option for the causal minimalist, because they apparently possess all the legally needed properties of consistency, simplicity and fairness sought by legal experts.

**Foreseeability and risk** Two other notions that have a probabilistic flavor may play an important role in the causal assessment of a case: the notions of foreseeability and risk. According to foreseeability theories, which are mostly adopted by common law jurisdictions, the defendant is liable for harm of which his conduct was a *sine qua non* condition only if the *type* of harm was objectively foreseeable. Moreover, foreseeability is not regarded as a test of causation in fact but as a requirement additional to it (i.e. a re-
quirement for legal causation).

The other (probabilistic) notion that is used in common law systems is risk. Hart and Honoré see it as a generalization of foreseeability (Hart and Honore, 1985) (p. 255). If defendant’s conduct is a *sine qua non* condition of harm, he is responsible if the harm falls within the risks to which his conduct has exposed others.

**Scope of the rule and equity** There exist other two adjunctive criteria, that are not causal, but that are proposed by the minimalists as complementary to counterfactuals or probabilities in attributing legal responsibility. These are the scope of the rule violated (minimalistic tests should be applied only if it is already clear that the considered harm is within the scope of the rule violated by the person whose conduct is under scrutiny) and equity (the minimalist seeks an imposition of liability that is equitable as between parties).

The most common criticism against causal minimalism concerns its tendency to sacrifice the complexity of most legal cases to the (apparent) simplicity (read it: simplism) of most logical and probabilistic criteria.

3.3. **Hart and Honoré’s solution**

The last approach to the problem of causation in fact that is relevant to the purpose of our research is Hart and Honoré’s approach. In their book (Hart and Honore, 1985) the two English authors argue against causal maximalism as well as against causal minimalism and they propose an alternative view, which is a synthesis of those two opposed approaches.

The arguments of the two scholars against the existing doctrines of legal-theoretical causation are somehow intertwined. We have already mentioned some of these arguments. We briefly summarize them here. On the one hand, Hart and Honoré see the notion of causal proximity as too vague for supporting a coherent form of legal analysis. Therefore, the definition of a clearer concept of causation in fact is needed. This should not become, though, an attempt at defining a set of rigid “rules for the determination of proximate cause”, because most of these rules for handling questions of fact (e.g. Beale’s) implicitly encode legal principles, i.e. answers to questions of policy. On the other hand, Hart and Honoré advocate that the correct usage of technical concepts such as foreseeability, risk and scope of the rule is highly dependent on (the application of) a plausible notion of causation in fact; and this notion, they say, is mostly hidden by causal minimalists in the *sine qua non* test (i.e., counterfactual reasoning), which is often proposed as the necessary
inferential counterpart of the technical legal notions mentioned above. In order to ease the tension between the common sense perception of the world and the (technical) legal view on it, Hart and Honoré propose a whole new approach to the problem. The main methodological novelty of their study is the idea of importing analytical philosophy into the discussion on legal causal reasoning (p. xxxiii, chapters I and II). The theoretical starting point of their analysis (p. xxxv) is the observation that the lack of agreement between causal maximalism and causal minimalism is due to a misunderstanding on whether it should be clause 2 or clause 3 of Definition 2 that prevails in the assignment of legal responsibility. While causal maximalism points everything on causation (clause 2), causal minimalism considers almost exclusively fault (clause 3). Neither of this two approaches, though, yields sufficiently convincing results because they both fail to provide an explicit account of the elements of a case that a judicial authority should consider when assessing causation in fact. Causal maximalism makes this mistake as a consequence of its excessively optimistic reliance on common sense and its capacity to find the right elements of causation every time this is needed. Causal minimalism makes this same mistake as a result of its excessively pessimistic view of common sense, as not being sufficiently logical.

According to Hart and Honoré the way out of this theoretical impasse may be found through an analytical approach. They propose to move from the classical common sense, logical or mathematical approaches to an analytical approach, which attempts at defining what it concretely means in legal settings for an agent to cause an event. There is some non logical knowledge of the world that must be taken into consideration when assessing causal relations. Hence, in order to find out what is such non logical generalization, the legal language of causation must be studied by means of ordinary language analysis, which, at the time when Hart and Honoré wrote, was a major development in the Anglo-Saxon philosophical landscape (development which is usually attributed to the “second” Wittgenstein or Austin). Such study should refrain from both temptations of explaining everything (as in causal maximalism) and of leaving everything unexplained (as in causal minimalism). The result of Hart and Honoré’s search may be spelled out in the following definition.

**DEFINITION 3 (Causation in fact).** *Agent A causes an event e, that might involve agent B, if either of the following holds:*

1. *A starts some physical process that leads to e;*

2. *A provides reasons or draws attention to reasons which influence the conduct of B, who causes e;*
3. A provides B with opportunities to cause e.

4. All the important negative variants of clauses 1, 2, 3

Now, an analysis of Example 2 according to Definition 3 yields a result that is not very different from those obtained by applying any maximalist or minimalist causal criteria. Either the mother or the seller are at the beginning of the causal chain. Contrary to most causal tests that we have seen so far, though, Definition 3 provides more clarity. It explicitly distinguishes the various types of causal links that connect each of the agents involved in to the harm. Rather than reducing every relevant causal relation to one type (e.g. proximate cause or \textit{sine qua non}), Definition 3 distinguishes various types of causal roles played by the persons involved in the case. For instance, the playmate (the boy who shot) has clearly started a physical process that leads to the harm (clause 1). The boy who owns the rifle has provided opportunities to the playmate for shooting. The mother of the rifle’s owner has failed to provide her child with reasons for not using the rifle and this has led him to play a causal role in the harm. Finally, the seller - if he has played any causal role in the harm at all - has provided an opportunity for the harm to come about.

Hart and Honoré’s approach tries to do exactly this: to make explicit what both the maximalist and the minimalist leave implicit. The simplicity of the tests built around proximity or counterfactuals hides a lot of complex assumptions. These, though, are implicitly employed by the legal expert in order to make sense of those tests. Hart and Honoré notice that such freedom of application jeopardizes the consistency of the tests, especially when employed over large corpora of cases. Therefore Definition 3 is proposed as a way of minimizing variations of outcome in legal causal analysis, by explicitly standardizing the implicit meaning of causal proximity or of counterfactual dependence.

Hart and Honoré’s effort at making things explicit is comparable under many respects to our own. For this reason in our treatment we adopt Definition 3 as our legal theoretical point of reference. In order to make it tractable for an \textit{automatic} interpreter, Definition 3 should be reconfigured. Such process of reconfiguration consists of three steps:

1. We first have to “open up” Definition 3, in order to have a better understanding of its internal elements and of their functioning.

2. Then, based on such understanding, we have to decide what parts of Definition 3 may be reduced to one another (or simply given up)

\footnote{An example of such negative variants is: A \textit{does not} provide reasons or \textit{does not} draw attention to reasons which might influence the conduct of B, who causes e.}
without losing too much of the overall original expressivity of the definition.

3. Finally, we may concentrate on reassembling the parts of Definition 3. In such final step, our focus is on gaining as much tractability as possible.

We proceed as follows. The next section is dedicated to a preliminary analysis of Hart and Honoré’s definition (opening up). Definition 3 distinguishes four different cases of causation in fact. In section 5, the cases of physical and agent causation is discussed from a philosophical perspective. Finally, in section 6 we will profit of this discussion and propose our own models of physical and agent causation as defined in the framework of CausatiOnt (reassembling).

4. Preliminary analysis

From an ontological viewpoint Definition 3 is redundant. Being constructed for the purpose of capturing a wide variety of cases of causation, the definition ranges over many of the levels of complexity characterizing our world. For instance, on the one hand, the concept of process is used for indicating some physical connection between an agent and a certain event; on the other hand, the concept of providing reasons is employed for indicating some psychological connection between two agents. From the definition it is not clear, though, what is the relation between a process and the providing of reasons. One might legitimately argue that providing reasons may be seen as a complex physical process that connects an agent to a (mental) event, which occurs within the (brain of) another agent (thus reducing clause 2 to clause 1). Another example of redundancy in Definition 3 is the use of concepts such as ‘starting’, ‘process’ or ‘event’. These concepts too might be defined in terms of one another.

Given the ontological ambiguities which characterize Definition 3, we want here to strive for some more clarity. In particular, we want to isolate all the terms that in the definition are particularly significant.

---

6 As explained in this section, providing reasons is a (causal) relation that holds between distinct agents (i.e., a case of interpersonal causation). Legally speaking, each agent’s beliefs, desires and intentions are part of the definition of the agent. Therefore, despite being reasons for action, it is implausible to admit that such beliefs, desires and intentions are provided by an agent to himself. It is only for the mentally insane that a distinction must be drawn between one’s beliefs, desires and intentions and one’s reasons for acting. But, then again, the mentally insane is usually not accountable, due to exactly such distinction.
from a causal perspective and group them according to the level of complexity to which they pertain (e.g., the physical level, the psychological level, etc.). Each of these levels comprises an ontologically distinct type of causation. In particular, Definition 3 refers to the following four main types of causation: physical causation, agent causation, interpersonal causation, negative causation.

**Physical causation** The case of physical causation is described by the final part of clause 1 of Definition 3, where the definition mentions a physical process that leads to an event. This is the intuitively most simple case of causation, in the sense that it is frequently quite simple to agree about cases of physical causation. For instance, in Example 2 it is rather obvious for anyone that the relation between the shooting of the gun and the wounding of the plaintiff is (physically speaking) a causal one. This is mainly due to the direct physical connection between the bullet and the victim’s eye.

**Agent causation** Agent causation is described by the initial part of clause 1 of Definition 3, where the definition mentions an agent starting a physical process. The agreement around cases of agent causation is not reached as easily as in cases of physical causation. This is due to the problem of detecting the beliefs, desires and intentions of the agent that starts the physical process. Usually a non intentional action does not fully qualify the agent that performs it as an agential cause, despite him being a physical cause. In Example 2, for instance, the boy who shoots clearly is a physical cause of the harm, simply because he pulled the trigger. His psychological state while pulling the trigger, though, can make a lot of difference on his status as the agential cause of the harm. Did he know what he was doing (i.e. pulling a trigger)? Did he intend or at least desire the consequent harm? Answers to this type of questions, even though not always positive ones, must be given (or assumed) in order to individuate agent causation in Example 2. Given the case description we provided, it is quite immediate to assume that the boy is an agentive cause of the harm. But what if it is shown that the boy inadvertently pulled the trigger while falling on the ground, after being pushed by the plaintiff himself? Or, again, what if it is shown that the boy acted under hypnosis or narcosis? In these cases the boy would not acquire the status of agential cause, despite holding a clear physical role in starting the harmful process.
Interpersonal causation

Things become even more complex when considering the role that interpersonal relations play in the chain of causation. One might be tempted to consider interpersonal causation just as a subcase of agent causation, where the psychological state of an agent exerts a causal influence on another agent. Things are not that simple, though. The causal influence that an agent may exercise on someone else may be physical in nature or psychological or a combination of the two. In an attempt at simplifying things, Definition 3 casts the intricate interpersonal causal relations in two separate clauses. Clause 2 refers to the psychological aspects of interpersonal causal relations as an agent giving another agent reasons or drawing another agent’s attention to reasons for causing something. Clause 3 refers to the physical aspects of causally relevant interactions as an agent providing another agent with opportunities for causing something. In other words, both reasons and opportunities act upon the psychological attitude of the agent that (physically) causes the harm. But, while reasons psychologically influence the attitude of an agent, opportunities physically sustain his attitude. This seems to be the main difference between the two cases of interpersonal causation individuated by Definition 3.

Negative causation

The most elusive case of causation is negative causation. Definition 3 refers to negative causation in clause 4 as all the important negative variants of the preceding clauses. Two intertwined problems are at stake here: an ontological problem and an epistemological problem. The ontological problem goes as follows. On the one hand, it is ontologically very difficult, almost paradoxical, to accept the general idea that something that does not exist can cause anything. This is quite evident in cases of physical causation: it does not make much sense, for instance, to say that the orbit of the Earth around the Sun is determined by Jupiter because this planet is not enough near to our planet to significantly divert its orbit. In other words, the principle applied to this astronomical example is that the absence of Jupiter can not determine the orbit of the Earth more than the absence of any other thing could. By applying this same principle to Example 1, one must say that the physical cause of the death of the traveler is not dehydration (i.e. the absence of water from the traveler’s body) but the (increasing) presence of heat. Physically speaking it is heating that causes the death of the desert traveler. Therefore, ontologically speaking, in order to see dehydration - rather than heating - as a cause, one must compare the actual
situation with an *ideal* or *optimal* condition of the traveler’s body. This type of comparison usually has some sense in cases of agent or interpersonal causation, where it is common to compare actual conducts with expected standards of behavior, because it is assumed that agents are free to determine their conduct - while physical objects are not. Therefore, negative causation is ontologically admissible in cases of agent and of interpersonal causation. Not all these types of cases admit negative versions, though. It is not simply a question of putting a negation before the positive formulation of one of the clauses. The choice of the admissible cases of negative agent or interpersonal causation poses interesting and (very) tough epistemological problems, which fall outside the scope of this article.

In the next section we present an overview of the main questions put by physical and agent causation. In our overall research we have not (yet) treated the problems put by interpersonal and negative causation, if not at an intuitive level. Therefore we do not provide any further information on such matters. The interested reader may refer to (Lehmann, 2003) (pp. 100-116) for a general introduction to these problems.

5. Philosophy on physical and agent causation

In this section we offer an overview of the main philosophical problems and approaches to physical and agent causation. This provides us with the necessary context for the definition of our own approach in the following section.

5.1. Physical causation

Physical causation is nowadays considered as the most general and comprehensive causal relation of all. Most of contemporary (western) culture works under the assumption that causal relations should be reduced to their (basic) physical aspects in order to be fully understood. For a simple and exhaustive historical overview of the development of the concept of causation, the reader is referred to the second chapter of (Hulswit, 1998). For a more easily accessible reference, the chapters about the Scientific Revolution of any handbook of Philosophy will do. A fairly entertaining one is, for instance, (Russell, 1984).

We provide here an overview of the formal properties, of the main approaches and of the ontological status of the causal relation. Moreover, from now on we only consider the very last part of Example 2, i.e. the
physical relations between the being pulled of the trigger (A), the being shot of the bullet (B) and the being hit of the plaintiff (C).

5.1.1. Formal properties of the causal relation
There are some formal properties of the causal relation which are a problem for any approach to (physical) causation. These are the classical logical relations of transitivity, symmetry and reflexivity. In particular one might ask:

1. Is causation a transitive relation? In other words, if A causes B and B causes C, does A cause C?

2. Is causation a symmetric relation? In other words, if A causes B, does B cause A?

3. Is causation a reflexive relation? Can A cause itself?

In our definition of physical causation (definition 17), we encapsulate the most widely accepted view that sees physical causation as a transitive, asymmetric and non reflexive relation.

5.1.2. Main approaches to the causal relation
There are a number of philosophical approaches to the problem of defining the right criteria for physical causation: logical, probabilistic, singularistic and functionalistic approaches, as follows.

Logical criteria The most traditional definition of causal relations is in terms of logical conditions. According to logical reductionism, the sentence A causes B can be thought as equivalent to one of the following logical statements: A is a necessary condition of B; or, A is a sufficient condition of B; or, A is a necessary and sufficient condition of B; or, A is an INUS condition of B, where the acronym INUS means: insufficient but non redundant part of an unnecessary but sufficient condition (Mackie, 1974); or, A is a counterfactual condition of B (Lewis, 1973).

Probabilistic criteria A second group of criteria that have been proposed for modeling and detecting causal relations are probabilistic criteria. Formally, the probabilistic relation that underlies a causal relation is expressed by the following inequality: \( p(B|A) > p(B|\neg A) \), which means that the conditional probability of B given A is higher than the conditional probability of B given not A.

Singularistic criteria There is a third group of criteria proposed in the philosophical literature on physical causation, which are known
as singularistic criteria. In very general terms, one may think of singularistic criteria as the conceptual counterpart of logical approaches based on counterfactuals. In other words, approaches based on the notion of change and logical approaches based on counterfactual dependencies may both be called singularistic. This is due to the fact that they both pivot on the principle that causation should not be defined in terms of any general principle or general causal law. As far as counterfactuals are concerned, such singularistic view leads to the definition of the truth conditions for causal statements in terms of counterfactual dependencies and in terms of a similarity relation between the actual world and the possible worlds accessible from it. The singularistic view on causation leads to the adoption of the rather neutral notion of change as the only parameter for detecting causation. In addition to change, spatio-temporal criteria usually play an important role too in accounts based on change. Thus, a singularist would for instance say that the cause of a particular change B is such particular change A as alone occurred in the immediate environment of B immediately before.

**Functionalist criteria** Functionalism is the last philosophic approach to causation that we want to mention here. This view on causation may be seen as the theoretical retreat of Philosophy under the practical advances of Science. To some extent, functionalism may be seen as the continuation of singularism by other means. The main difference from singularism is that functionalism seeks sharper tools than the notions of change or of counterfactual dependence for detecting causation. The various functionalist views proposed so far ((Russell, 1912), (Russell, 1948), (Salmon, 1984), (Dowe, 1995)) try to reduce the notion of causation to physical notions such as energy or momentum transfer between processes, in accordance to contemporary Physics. For instance a functionalist would consider a relation between A and B as causal, if the actual physical intersection between A and B involves exchange of a conserved quantity (e.g. energy).

The approach that is most suitable to our purposes is a combination of singularism and functionalism, because these two approaches are the most explicit ones in defining the physical elements of physical causation. In section 6 we explain how singularism and functionalism converge on our own view.

---

7 The momentum of an object is its propensity to continue moving, because of its mass and speed.
5.1.3. *The ontological status of the causal relata*

In addition to the questions concerning what a causal relation is and how can it be best described, there is the question regarding the ontological nature of the relata of a causal relation. There are three main options here: objects, states of affairs or events. The causal relata usually adopted in the (singularistic) literature on causation are events, because these bring about change. Moreover, it is easier to include objects and states of affairs in the definition of events, rather than the other way around; and events undoubtedly have a tighter conceptual relation with time, which in turn is deeply related to the notion of change. One of the main consequences of adopting events, is that all other ways of expressing the causal relation (i.e. as a connection between objects and/or states of affairs) must be considered only as linguistic tricks, i.e. shortcuts taken by the layman in order to avoid long, albeit proper, explanations in terms of events.

5.2. *Agent causation*

As explained in the introductory part of the subsection on physical causation, the general shift to physical reductionism, which started with the Scientific Revolution, has apparently left very small theoretical space for other forms of causation. If causation is only a physical sequence of causes and effects, there is no way to account for a whole range of mental phenomena that do not seem to be (exclusively) physical but that do anyway (seem to) have causal influence on the world. We provide here an overview of the main problems faced and some solution proposed by philosophers of actions.

There are two basic philosophical problems that concern action theoreticians who deal with agent causation:

**The nature of action and agency** This problem has given rise to various classifications of actions along two main dimensions: the knowledge of one’s own actions (awareness) and the control on one’s own action (governance). Depending on the awareness and on the control of the actor, five main types of actions may be defined: unconscious behavior, goal directed activity, involuntary behavior, intentional action and, finally, autonomous action. In their turn, these five types of actions raise tough philosophical problems concerning the actor’s governance over the movements of his body and over the causal consequences of such movements. One might think that the degree of control is directly proportional to the degree of awareness. This is not always true though, especially in the case of interaction with the physical world. Philosophers have therefore limited the notion of control to the body movements of the actor.
and they have explained such control either in terms of *proximity* (between the actor’s intentions and his motoric system) or in terms of *trying* (which should play the role of common denominator of all actions).

**The explanation of actions** The philosophical positions on this matter are two. On the one hand, Davidson and other theorists defend the position that reason explanations are causal explanations, which refer to the agent’s desires, intentions, and beliefs. On the other hand, so-called non-causalists, maintain that motivating reasons cannot be causes of action. Non-causalists stick to the principle of explanatory exclusion which affirms that either the reason explanations or the neural explanations are incorrect.

In the ontology proposed in the next section we adopt the notion of trying as the distinguishing factor between actions and mere body movements. Furthermore we adopt a Davidsonian view on reason explanations, as we find it much more natural to explain actions in terms of motives, reasons and other (so-called) mental entities (e.g., intentions and believes).

6. **CAUSATI Ont**

This section presents some parts of CausatiOnt, the parts that contain the concepts that are minimally needed for defining the relations of physical and agent causation. For conciseness we do not spend much time on matters of representation. Our ontology was implemented in Protégé-2000, a fairly liberal knowledge representation tool, based on the classical is-a relation. The figures, shown in the following, reproduce part of our is-a hierarchy. It should be noticed that such figures were automatically generated by Ontoviz, the graphic tool of Protégé-2000. For this reason they may sometimes be visually confusing, in the sense that they do not always mirror the order of presentation of our ontology in this article. We, therefore, have to ask for some patience from the reader, for the purpose of being conducted to a correct reading of the figures.

6.1. **CAUSATI Ont**’s top level

We present here the contents of the four top boxes shown in figure 1: noesis, category, dimension and entity. We illustrate both the intended meaning of such terms and we briefly introduce the subsumed concepts that are relevant to the present exposition.
6.1.1. **Noesis**

The first ontological choice, to which we commit ourselves, is avoiding to make a truly ontological choice. After years of struggling with the notion of causal relation (and those related to it, e.g. matter, object, process, energy, work, power, etc.) two main general properties of causal relations appeared quite clear:

1. **Causal relations are neither purely ontological nor purely epistemological.** They are something in between. A child, for instance, learns that fire is harmful by touching it and experiencing the painful sensations that it causes. This hard ontological fact is at the basis of the (subsequent) epistemological generalization that brings the adult to be aware of the harm caused by hot things in general. Similarly to the causal awareness of individuals, the causal awareness of humanity is developmental. We have mainly learnt about causal relations by a process of trial and error.

2. **Therefore, the representation of causal knowledge cannot be limited to the ontological elements of causal relations (i.e. the entities).** It must be extended to the epistemological elements (i.e. the categories) and to the phenomenological relations between them (i.e. the dimensions). This extension might seem as a non parsimonious scientific practice, and it probably is. But it gives us some room to explain what in causal reasoning pertains to us as observing entities and what pertains to the world as observed entity. Furthermore, by not limiting ourselves to ontology we provide a clear way of distinguishing semantically similar terms (e.g., matter, a category; mass, a dimension; object, an entity) within the same AI-like ontology.

Given 1. and 2. we define the top class of our ontology as follows.

**DEFINITION 4** (Noesis). *Noesis is the psychological counterpart of experience (i.e. perception, learning and reasoning).*
The notion of noesis has a rather long philosophical tradition, which dates back to Greek Philosophy. A number of words stemming from the root *nous-* (e.g. *noein*, thinking, *nous*, mind, *noumenon*, the object of thought) were used to indicate whatever was related to thought as a psychological activity. In the XX century, the notions of noesis and noemata were employed by the phenomenologists (e.g. Husserl) to indicate, respectively, the process and the result of perceptual and intellectual activity.

As far as we are concerned, we adopt here the notion of noesis in its broadest cognitive sense. We consider all the experiences of an individual human being to be physical phenomena. On the one hand, perceptual experiences (e.g. perceiving the color red) are the result of the interaction between the physical world (i.e. light) and an individual’s sensory system (e.g. his optic nerve and other parts of his brain). On the other hand, intellectual experiences (e.g. thinking about the notion of color) occur in the brain, i.e. they too are physical phenomena. Besides their physical nature, though, both perceptual and intellectual experiences generally seem to have a psychological counterpart, i.e. a part of which the individual is aware (i.e. the color red, in the example of perceptual experiences, and the notion of color, in the example of intellectual experiences). Any such psychological counterpart of an experience is noesis.

6.1.2. *Category*

A category is that part of noesis, which cannot be (philosophically) reduced to any other parts. It must therefore be assumed as a basic intellectual element\(^8\), which structures our perceptual experience of the world and our reasoning about the world. We define it as follows.

**DEFINITION 5 (Category).** *Category is knowledge-related (i.e. epistemological) noesis.*

Categories form the intellectual background of our noetic experience of the world (i.e. of our perception, learning and reasoning about the world). Even though categories play a crucial role in noesis, we are hardly aware of them in our experience. When perceiving, learning or reasoning we are not fully aware of the categories that are supporting our effort. For instance, when reasoning about (i.e. having an intellectual experience of) or perceiving (i.e. having a physical experience of) an entity (e.g. an object), a number of categories (e.g. matter and quantity) make our experience possible, even though they are not immediately present to our mind and/or to our sensory system.

---

\(^8\) The expression ‘assuming a notion as basic intellectual element’ is used here as a synonym of ‘postulating a notion’. In other words, categories must be postulated.
Categories are, therefore, here understood as in (Kantian) Epistemology: as the basic notions on which our (intellectual and perceptual) experience builds up. Categories are different from sets, classes and instances. They are different from sets because they are not collections of (prototypical) individuals. Categories are different from classes because they have no properties describing them. Finally, categories are different from (prototypical) individuals because they are universal not particular.

We consider it to be important for an AI-like ontology to contain a description of the categories on which it is based. Even though categories do not play a direct role in reasoning about and/or perceiving entities, they may play two other important roles in an ontology:

1. On the one hand, as support to reasoning about knowledge at a general level.

2. On the other hand, as purely descriptive notions that clarify the intuitive meaning of the terms that are used in reasoning about entities (which we call the dimensions, see below).

Our intent is to use categories in the sense indicated at point 2 above. We therefore present a number of terms that indicate the basic epistemological structure assumed in our ontology. As shown in figure 2 we distinguish between two main groups of categories: the categories of existence and the categories of experience. The opposition between these two types of categories is the epistemological equivalent of the opposition, within noesis, between entity (or Ontology) and category (or Epistemology). In other words, just like in noesis, where we distinguish existence (the entity) from knowledge (the category), in category we distinguish between the knowledge of what exists (category of existence) from the knowledge of the modes of knowledge (category of experience). These second categories describe how we know what exists (or, rather, how we know the categories of existence). Categories of existence encompass notions such as space, matter, energy, change and others; whereas category of experience encompass notions such as quantity, quality and time.

An important subcategory of change is causality. There has been a lot of philosophical discussion concerning both the ontological and the epistemological status of causal relations. In accordance with (Hulswit, 1998) we provide here the first half of the solution proposed by us. We

---

9 We want to avoid to use here the expression *a priori* in order to describe the status of categories. As a matter of fact, under a noetical perspective nothing is *a priori* and one may see categories as the result of evolution, both of individuals and of species.
adopt the term ‘causality’ to indicate the epistemological counterpart of an ontological dependence. We see here causality as the epistemological counterpart of an ontological dependence between categories of existence. In other words, from an epistemological point of view causality is an ontological constraint between the categories of existence. Such constraint fixes the possibility of existence of one category relative to another. We propose here to adopt the following ontological dependence between categories of existence as standard notion of common sense causality: will cannot exist without life, life cannot exist without energy, matter and space. In other words there is an ontological dependence between will and life, on the one hand, and energy, matter and space, on the other hand.

6.1.3. Dimension
From the (theoretical) background consisting of the categories the dimensions emerge, which we define as follows.

DEFINITION 6 (Dimension). Dimension is experience-related (i.e. phenomenological) noesis. A dimension relates two categories.

The standard example of a dimension is mass. By experience, all physical objects have a mass, which is the quantity of matter they comprise. We never have, though, a concrete experience of either matter or quantity as such. Therefore, we must assume their existence as categories, rather than as entities, and employ them in the definition of the notion of mass. In other words, the concrete notion of mass relates the epistemological to the ontological part of our noetic experience. We experience objects (ontology) as having mass (phenomenology), which
relates two categories: matter and quantity (epistemology).
In the definitions of dimensions, we associate categories to one another
with the expression ‘experienced by means of’. This is to underline
the fact that the definition of dimensions in terms of categories is not
an ontological but a phenomenological definition. We therefore say, for
instance, that mass is matter \textit{experienced by means of} quantity (rather
than mass \textit{is} a quantity of matter), where the experience of matter
by means of quantity is a purely \textit{intellectual} one, as both matter and
quantity are categories, not entities.
Furthermore, it should be noticed that we use the expression ‘ex-
perienced by means of’ also in the definition of entities in terms of
dimensions. In this case, the expression ‘experienced by means of’ refers
to the \textit{perceptual} (rather than the intellectual) experience of an entity
(e.g. an object) through a dimension (e.g. mass).
The following dimensions have been defined: volume (space experienced
by means of quantity), form (space experienced by means of quality),
location (space experienced by means of time), mass (matter experi-
exenced by means of quantity), material (matter experienced by means
of quality), state (matter experienced by means of time), work (energy
experienced by means of quantity), energy-form (energy experienced
by means of quality), power (energy experienced by means of time),
direction (change experienced by means of quantity), transition (change
experienced by means of quality), period (change experienced by means
of time), sensitivity (life experienced by means of quantity), instinct
(life experienced by means of quality), age (life experienced by means
of time), intentionality (will experienced by means of quantity), repre-
sentational content (will experienced by means of quality), enactment
(will experienced by means of time).
Some of such dimensions are shown in figures 3 and 4, as descriptors of
a class (for instance, dimension ‘mass’ is a descriptor of class ‘object’
and usually takes numeric values).

6.1.4. \textit{Entity}
The notion of entity indicates something that exists separately from
other things and has a clear identity. Entities are the subject of study
of ontology. We define entity as follows.

\textbf{DEFINITION 7 (Entity).} \textit{Entity is existence-related (i.e. ontological)}
\textit{noesis}.

In Example 2 examples of (different types of) entities are: the event
of the trigger being pulled; the boy who pulls the trigger (both his
body and his mind); the process of being pulled; the trigger; the causal
connection between the event of the trigger being pulled and the event of the plaintiff being hit.

6.2. **CAUSATI**$^N$T’s CAUSAL ENTITIES

By means of the dimensions we define here some of the entities needed for the definition of causal relations. We proceed as follows: we firstly introduce the definition of physical entity, object and process; we then define the notion of mental entity; and, finally, we define the notions of occurrence and event.

The dimensions that stem from the categories of physical existence (i.e. space, matter, energy, change) provide us with the necessary terminology for defining physical entities. We define physical entity as follows.

DEFINITION 8 (Physical entity). Physical entity is an entity experienced by means of one or more of the following dimensions: volume, form, location, mass, material, state, work, energy-form, power, direction, transition, period, sensitivity, instinct, age.

![Figure 3. CAUSATI$^N$T’s class Physical Entity and subclasses](image)

As shown in figure 3, there are three main physical entities: objects, organisms (which we do not treat here) and processes. They are defined as follows.

DEFINITION 9 (Object). Object is a physical entity which is experienced by means of all of the following dimensions: volume, form, location, mass, material, state.

In Example 2 an example of object is the trigger.
DEFINITION 10 (Process). Process is a physical entity experienced by means of all of the following dimensions: work, energy-form, power, direction, transition, period.

In Example 2 an example of process is being pulled. The dimensions that stem from the categories of psychological existence (i.e. will) provide us with the necessary terminology for defining mental entities.

DEFINITION 11 (Mental entity). Mental entity is either:

1. An entity experienced by means of one or more of the following dimensions: intentionality, representational content and enactment.

2. Or a process that has either of the following dimensions as transition: intentionality, representational content and enactment.

As shown in figure 4 there are two main mental entities: thought and mental process. Thought is often understood as the content of some cognitive state. Just as for the case of the notion of object we are able to provide here a modular definition of the notion of thought, as follows.

DEFINITION 12 (Thought). Thought is a mental entity which is experienced by means of all of the following dimensions: intentionality, representational content and enactment.

In example 2 a thought is the boy’s desire to buy the air rifle. Mental processes form a special type of processes. On the one hand, they are described by the same dimensions as physical and biological processes (i.e. work, energy-form, power, direction, main transition, period): mental processes actually are physical processes, in the sense...
that they (must) have some physical counterpart. On the other hand, though, they are also mental entities, because their transitions pertain to the dimensions of thoughts rather than of objects or organisms. In other words, mental processes unify the types of entities that are generally considered as characterizing the notion of agency: physical (in its broad sense, including biological) and mental entities.

**DEFINITION 13** (Mental process). *Mental process is both a mental entity and a physical process which has either of the following dimensions as its transition: intentionality, representational content and enactment.*

In example 2 a mental process is the action of buying the air rifle. Another type of entity, non physical but relational, must be introduced here in order to subsequently generate the notion of event.

**DEFINITION 14** (Occurrence). *Occurrence is a reified relation between objects, processes, thoughts and/or occurrences.*

An event is a reified relation between a process, an object, an action and an agent. We define an event as follows.

**DEFINITION 15** (Event). *Event is an occurrence of a process (the occurrence) involving an object (the subject), where the process is possibly initiated by an action (the act) of an agent (the actor).*

In Example 2 an example of event is the trigger being pulled. Finally, the notion of causation may be defined.

**DEFINITION 16** (Causation). *Causation is an occurrence of two events, the cause and the effect.*

Two remarks are needed here. Firstly, the relation of causation introduced in Definition 16 is the counterpart, within our ontology, of the relation introduced in Definition 3 (i.e., causation in fact). It should be noticed that Definition 16 subsumes Definition 3, as the former can subsume more types of causation than those subsumed by causation in fact (which are, according to the preliminary analysis of section 4, physical, agent, interpersonal and negative causation). Definitions 17 and 18 define the first two of such types. Secondly, Definition 16 is very broad and it is needed as a definitional node in the ontology. In other words, all the clauses that provide the sufficient conditions for more restrictive (and therefore more interesting) causal relations are provided in the definitions subsumed by Definition 16. This does not mean that the relation introduced in Definition 16 is indistinguishable
from simple sequencing of events. It should not be forgotten that Definition 16 introduces a type of occurrence. This has, of course, a rather strong implication: by definition all reified relations between events are causal relations. But, on the other hand, a niche is carved for logical (i.e. non reified) relations between events (e.g. temporal or mereological relations), which keep their non causal nature.

Physical causation is a type of causation and there are a number of conditions that must be met by any two considered events in order for physical causation to hold.

DEFINITION 17 (Physical causation). Physical causation is causation between an event $E_1$, which is an occurrence of a physical process $P_1$ (the occurrence of) involving an object $O_1$ (the subject), and event $E_2$, which is an occurrence of a physical process $P_2$ (the occurrence of) involving an object $O_2$ (the subject). A relation of physical causation holds between $E_1$, the cause, and $E_2$, the effect, if the following conditions are met:

1. $O_1$ and $O_2$ are not the same object, according to the adopted identity criterion for objects.
   Comment: the subjects must be truly distinguished objects.

2. $P_1$ and $P_2$ are not the same process, according to the adopted identity criterion for processes.
   Comment: an event cannot cause itself. By this clause we adopt the view that causation is an irreflexive relation.

3. $P_1$’s period precedes $P_2$’s period.
   Comment: the cause temporally precedes the effect. Even for processes that are temporally distributed (i.e. continuous) the causing process starts before the caused one. By this clause we adopt the view that causation is a temporally asymmetric relation.

4. $P_1$’s energy-form is the same as $P_2$’s energy-form or $E_2$ is reducible to events $E_{2,1} \ldots E_{2,n}$ such that:
   
   a) $E_{2,1} \ldots E_{2,n}$ are occurrences of processes $P_{2,1} \ldots P_{2,n}$, which all have the same energy form of $P_1$.
   
   b) $E_{2,1} \ldots E_{2,n}$ have as their subjects objects $O_{2,1} \ldots O_{2,n}$, which are the grains of $O_2$, according to the adopted structural constraints.

In the interaction between two objects energy is transferred or transformed. In this latter case, the transformation of energy should be reducible to a transfer of energy between the cause and the events occurring to the grains of the object of the effect.
5. \( P_1 \)'s direction is the same as \( P_2 \)'s direction or \( P_1 \)'s power is greater or equal to \( P_2 \)'s power or \( P_1 \)'s work is greater or equal to \( P_2 \)'s work.

Comment: this clause accounts for the fact that usually changes of one sign cause changes of the same sign (i.e. an increase can usually only be caused by an increase and a decrease by a decrease). If this condition cannot be tested (which might be the case when lack of information makes it impossible to establish the directions of either \( P_1 \) or \( P_2 \)) or if it is not satisfied, one may want to use the principle of the dispersion of energy in order to distinguish the cause from the effect.

6. The category of existence of \( P_2 \)'s transition can not exist without the category of existence of \( P_1 \)'s transition, according to the adopted causality constraint.

Comment: changes in \( O_1 \)'s dimensions can only affect those dimensions of \( O_2 \) that are ontologically dependent on the dimensions changed in \( O_1 \), according to the adopted causality.

It should be added that we take physical causation to be a transitive relation. In Example 2 an example of physical causation is the relation between the event of the trigger being pulled and the event of the plaintiff being hit.

As far as agent causation is concerned we take it to be a subclass of physical causation. There are a number of conditions that must be met by the two considered events in order for agent causation to hold. The most prominent of such conditions are the existence of a relation of physical causation between the first and the second event and the knowledgeability of the actor of the first event.

DEFINITION 18 (Agent causation). Agent causation is causation between an event \( E_1 \), which is the occurrence of a physical process \( P_1 \), initiated by actor \( Ag_1 \) by act \( Ac_1 \), to a subject \( O_1 \) and event \( E_2 \), which is the occurrence of physical process \( P_2 \) to subject \( O_2 \). A relation of agent causation holds between \( E_1 \), the cause, and \( E_2 \), the effect, if the following conditions are met:

1. \( E_2 \) is not initiated by any actor, say, \( Ag_2 \) performing any act, say, \( Ac_2 \).

Comment: There is no novus actus interveniens between the cause and the effect.

2. There is a relation of physical causation between \( E_1 \) and \( E_2 \).

3. \( E_2 \) is the representational content of a thought with no enactment that is in \( Ag_1 \)'s mind.
Comment: The actor must have in mind the result of the process (directly or indirectly initiated by his action. Such awareness can range from simple belief (i.e. foreseeability) to intent (goal).

Finally, we consider agent causation to be an intransitive relation. This property of the relation descends from clause 1 of definition 18.

7. Conclusion

Reasoning about causation in fact is an essential element of attributing legal responsibility. Therefore, the automation of the attribution of legal responsibility requires a modelling effort aimed at the following: a thorough understanding of the relation between the legal concepts of responsibility and of causation in fact; a thorough understanding of the relation between causation in fact and the common sense concept of causation; and, finally, the specification of an ontology of the concepts that are minimally required for (automatic) common sense reasoning about causation.

In this article we offered a worked out example of the indicated analysis. Such example consists of: a definition of the legal concept of responsibility (in terms of liability and accountability); a definition of the legal concept of causation in fact (in terms of the initiation of physical processes by an agent and of the provision of reasons and/or opportunities to other agents); an AI-like ontology, called CausatiOnt, of the common sense (causal) concepts that are minimally needed for reasoning about the legal concept of causation in fact (in particular, the concepts of category, dimension, object, agent, process, event and act).

References