DOI:
10.1145/2746090.2746102

Document Version
Early version, also known as pre-print

Causal Analysis for Attributing Responsibility in Legal Cases

Hana Chockler
Department of Informatics
King's College London, UK.
hana.chockler@kcl.ac.uk

Norman Fenton
School of Electronic Engineering and Computer Science
Queen Mary University of London, UK
n.fenton@qmul.ac.uk

Jeroen Keppens
Department of Informatics
King's College London, UK.
jeroen.keppens@kcl.ac.uk

David A. Lagnado
Department of Experimental Psychology
University College London, UK
d.lagnado@ucl.ac.uk

ABSTRACT

An important challenge in the field of law is the attribution of responsibility and blame to individuals and organisations for a given harm. Attributing legal responsibility often involves (but is not limited to) assessing to what extent certain parties have caused harm, or could have prevented harm from occurring. This paper presents a causal framework for performing such assessments that is particularly suitable for the analysis of complex legal cases, where the actions of many parties have had a direct or indirect effect on the harm that did occur. This framework is evaluated by means of a case study that applies it to the Baby P. case, a high-profile case of child abuse leading to the death of a child that has been the subject of a number of public inquiries in the UK. The paper concludes with a discussion of the framework, including a roadmap of future work and barriers to adoption.

Categories and Subject Descriptors

H.4.2 [**Information Systems Applications**]: Types of systems—*decision support*; J.1 [**Computer Applications**]: Administrative Data Processing—*Law*; J.4 [**Computer Applications**]: Social and Behavioral Sciences—*Psychology, Sociology*

Keywords

Causality, Responsibility, Legal cases, Inquest, Enquiry, Quantitative analysis of responsibility, Blame

ICAIL'15 June 08 - 12, 2015, San Diego, CA, USA

http://dx.doi.org/10.1145/2746090.2746102 ...$15.00.

1. INTRODUCTION

Responsibility attribution is pervasive in many areas of human interaction, including law, politics, business, sports, and everyday social settings. Something goes wrong, and we seek to assign blame to someone or some group. This is a challenging process: there are often complex causal interactions between multiple agents and events, and agents can have differing roles and varying degrees of involvement. One needs to work out what happened, who did what, why people acted as they did, and what would have happened if they had acted differently (e.g. would the same results have been obtained?).

The problem of responsibility attribution looms especially large in the law. Legal responsibility is a complex concept that depends on a variety of factors [30], including a person's behaviour, the lawfulness of this behaviour (e.g., did it break a duty or contract?), whether the behaviour caused the harm at issue, and the reasons for this behaviour. Causation is often a central question: Did the defendant's action cause or contribute to the result? Did the actions of a third party intervene to cut the causal chain from the defendant's action to the final outcome? How did the two agents' actions combine to yield the result [22]?

Despite the key role of causation, the law lacks a comprehensive theoretical or formal framework for causal analysis. The two most common accounts, the "but-for" and NESS tests [35] struggle to accommodate complex networks of interacting causes, overdetermination and pre-emption, probabilistic causes, and foreseeability [13, 33]. More recent legal accounts, for example [33], do capture some of these issues in an informal way, but without a framework that permits formalization and objective assessment. The causal framework presented in this paper seeks to address these issues. It provides a formal account that extends the logic of but-for causation to complex interacting networks of causes, solving problems of *overdetermination* and *pre-emption*, and capturing notions of *foreseeability* and *probabilistic causes*. It relies on the extension of formal definitions of responsibility and blame, and supplies a metric for assigning degrees of responsibility and blame – especially suitable for modelling complex situations with multiple agents. Moreover, recent psychological studies suggest that people's intuitive responsibility judgements are modelled accurately by the framework [15, 36, 28].

This rich formal framework permits causal analysis of complex

legal cases. It takes into account the various factors that need to be proved legally, and thus allows for more principled assessment of legal cases, as demonstrated by the case study in the paper.

2. DOMAINS OF APPLICATION

If an offence (either criminal or civil) is subject to any kind of legal investigation, the ultimate objective of that investigation is to attribute responsibility for the offence and to determine appropriate punishment for the offender(s). For many offences there are multiple people and/or organisations that bear some degree of responsibility, and the ultimate punishment is expected to take account of this. For example:

1. In a criminal case involving the death of a child (such as the well-known cases of Victoria Climbié [2] and Baby P. [21]) the trial may determine that the baby died as a result of being struck by the father, while the mother deliberately failed to report the incident to the authorities. Although both parents bear responsibility for the death (it is assumed the child may have survived if treated quickly) it is likely they will receive different sentences to reflect the different levels of responsibility.

2. An inquest or inquiry, following a case like 1, may result in a much broader investigation. For example, in both of the actual cases mentioned there, subsequent public inquiries covered the actions of the social services department, local health authority, police and specific individuals within those organisations. Here, the objective is to attribute responsibility so that appropriate action can be taken. If an organisation or individual is found to bear a "high responsibility" then further charges and/or punishment may follow, while for lower levels of responsibility it may be sufficient to make recommendations for behaviour/policy change to prevent similar problems occurring in the future.

3. Complex civil cases often involve a claimant suing multiple parties. For example, there are numerous legal cases where a victim of a work-related illness (e.g., lung cancer) claims against several different employees (e.g., [1]; more generally, see [33, 16]). Medical negligence also provides complex cases with multiple agents involved in harm to the claimant. If and when a final sum is agreed, the amount to be paid by each party should be proportional to their level of responsibility.

4. An enquiry such as that in 2 can lead to even further *removed* levels of responsibility in subsequent cases. For example, after the doctor Harold Shipman was found guilty of murdering multiple patients the subsequent inquiry attributed blame to a large number of named medical professionals for failure to understand and act on what was happening [3]. As a result of this, there are now widespread reports of what is referred to as the "Shipman effect" in cases involving "suspicious" patient deaths, whereby medical professionals may have been over-zealous in reporting colleagues. This has led to potential miscarriages of justice for which those over-zealous reporters have some responsibility.

This paper focuses on such situations, with the objective of providing a formal framework to attribute responsibility fairly. The framework complements other recent work in legal reasoning that deals, for example, with causal explanatory Bayesian networks for legal reasoning and probabilistic evaluation of evidence [12, 29, 34], and more generic models of legal argumentation [7]. That work can be considered most relevant at (i) the evidence gathering stage [11, 34] (for example, to help the Crown Prosecution Service determine whether there is sufficient evidence for a likely conviction) and (ii) in helping lawyers understand and present evidence. However, the work in this paper is most obviously targeted at later stages in the legal process (such as in both sentencing and inquests) - specifically when there is less uncertainty about the evidence and more of a focus on attributing responsibility.

3. CAUSALITY, RESPONSIBILITY, BLAME, AND NORMALITY

In this section, we review Halpern and Pearl's definitions of causal models and causality [20], and Chockler and Halpern's definitions of responsibility and blame [8] and show how these definitions are adapted to legal settings.

3.1 Causal models

Formally, a causal model M is a tuple $\langle \mathcal{S}, \mathcal{F} \rangle$, where $\mathcal{S}$ is a set of variables, associating with each variable the range of its values, and $\mathcal{F}$ is a set of functions defining the dependencies between the variables. The set of variables is usually partitioned into two subsets: the exogenous variables $\mathcal{U}$, whose values are determined by the factors outside the model, and the endogenous variables $\mathcal{V}$, whose values are determined by the exogenous variables and the functions $\mathcal{F}$.

We can describe a causal model M using a *causal network*, which is, roughly speaking, a directed acyclic graph representing variables and their dependencies. We call a setting $\vec{u}$ for the variables in $\mathcal{U}$ a *context*. The equations determined by $\mathcal{F} = \{F_X : X \in \mathcal{V}\}$ can be thought of as representing processes (or mechanisms) by which values are assigned to variables. For example, if $F_X(Y, Z, U) = Y + U$ (which we usually write as $X = Y + U$), then if $Y = 3$ and $U = 2$, then $X = 5$, regardless of how Z is set.

While the equations for a given problem are typically obvious, the choice of variables may not be. Consider the following example (due to Hall [18]), showing that the choice of variables influences the causal analysis. Suppose that Suzy and Billy both pick up rocks and throw them at a bottle. Suzy's rock gets there first, shattering the bottle. Since both throws are perfectly accurate, Billy's would have shattered the bottle had Suzy not thrown. In this case, a naive model might have an exogenous variable U that encapsulates whatever background factors cause Suzy and Billy to decide to throw the rock (the details of U do not matter, since we are interested only in the context where U's value is such that both Suzy and Billy throw), a variable *ST* for Suzy throws (*ST* $= 1$ if Suzy throws, and $ST = 0$ if she doesn't), a variable *BT* for Billy throws, and a variable *BS* for bottle shatters. In the naive model, whose graph is given in Figure 1 on the left, BS is 1 if one of *ST* and *BT* is 1. This causal model does

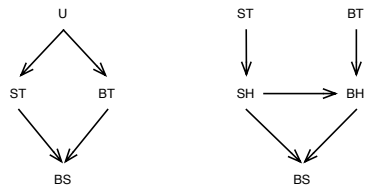

Figure 1: Models for the rock-throwing example.

not distinguish between Suzy and Billy's rocks hitting the bottle simultaneously and Suzy's rock hitting first. It also illustrates the concept of *overdetermination*, mentioned in the introduction, with the bottle shattering caused by both rocks.

A more sophisticated model might also include variables *SH* and *BH*, for Suzy's rock hits the bottle and Billy's rock hits the bottle. Clearly *BS* is 1 iff one of *SH* and *BH* is 1. However, now, *SH* is 1 if *ST* is 1, and $BH = 1$ if $BT = 1$ and $SH = 0$. Thus, Billy's throw hits if Billy throws *and* Suzy's rock doesn't hit, capturing *pre-emption* of Billy's rock by Suzy's rock. This model is described by the graph on the right in Figure 1, where we implicitly assume a context where Suzy throws first, so there is an edge from *SH* to *BH*, but not one in the other direction.[1]

Given a causal model $M = (\mathcal{S}, \mathcal{F})$, a (possibly empty) vector $\vec{X}$ of variables in $\mathcal{V}$, and a vector $\vec{x}$ of values for the variables in $\vec{X}$, we define a new causal model, denoted $M_{\vec{X} \leftarrow \vec{x}}$, which is identical to M, except that the equation for the variables $\vec{X}$ in $\mathcal{F}$ is replaced by $\vec{X} = \vec{x}$. Intuitively, this is the causal model that results when the variables in $\vec{X}$ are set to $\vec{x}$ by some external action that affects only the variables in $\vec{X}$ (and overrides the effects of the causal equations). For example, if M is the more sophisticated model for the rock-throwing example, then $M_{ST \leftarrow 0}$ is the model where Suzy doesn't throw.

A causal (propositional) formula φ is true or false in a causal model, given a *context*. We write $(M, \vec{u}) \models \varphi$ if φ is true in causal model M given context $\vec{u}$. $(M, \vec{u}) \models [\vec{Y} \leftarrow \vec{y}](X = x)$ if the variable X has value x in the unique (since we are dealing with recursive models) solution to the equations in $M_{\vec{Y} \leftarrow \vec{y}}$ in context $\vec{u}$ (i.e., the unique vector of values for the exogenous variables that simultaneously satisfies all equations $F_Z^{\vec{Y} \leftarrow \vec{y}}$, $Z \in \mathcal{V} - \vec{Y}$, with the variables in $\mathcal{U}$ set to $\vec{u}$). These definitions can be extended to arbitrary causal formulas.

3.2 Causality

We now review the definition of causality by Halpern and Pearl [20].

DEFINITION 3.1. *$\vec{X} = \vec{x}$ is a* cause *of φ in $(M, \vec{u})$ if the following three conditions hold:*

AC1. $(M, \vec{u}) \models (\vec{X} = \vec{x}) \wedge \varphi$.

AC2. *There exist a partition $(\vec{Z}, \vec{W})$ of $\mathcal{V}$ with $\vec{X} \subseteq \vec{Z}$ and some setting $(\vec{x}', \vec{w})$ of the variables in $(\vec{X}, \vec{W})$ such that if $(M, \vec{u}) \models Z = z^*$ for $Z \in \vec{Z}$, then*

- *(a)* $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}]\neg\varphi$.
- *(b)* $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}, \vec{W}' \leftarrow \vec{w}, \vec{Z}' \leftarrow \vec{z}^*]\varphi$ *for all subsets $\vec{Z}'$ of $\vec{Z} \setminus \vec{X}$ and all subsets $\vec{W}'$ of $\vec{W}$, where we abuse notation and write $\vec{W}' \leftarrow \vec{w}$ to denote the assignment where the variables in $\vec{W}'$ get the same values as they would in the assignment $\vec{W} \leftarrow \vec{w}$, and similarly for $\vec{Z}' \leftarrow \vec{z}^*$. That is, setting any subset $\vec{W}'$ of $\vec{W}$ to the values in $\vec{w}$ should have no effect on φ as long as $\vec{X}$ has the value $\vec{x}$, even if all the variables in an arbitrary subset of $\vec{Z}$ are set to their original values in the context $\vec{u}$.*

AC3. *$(\vec{X} = \vec{x})$ is minimal; no subset of $\vec{X}$ satisfies AC2.*

[1]Note that, strictly speaking, we do not need a separate variable for SH, as in our equations, $SH = ST$, and therefore, we can just use ST everywhere; we add it here to explicitly capture the order of rocks hitting the bottle.

If $\vec{X}$ is a singleton, then $X = x$ is said to be a *singleton cause* of φ in $(M, \vec{u})$.

Admittedly, the considerations that led to this definition are quite subtle. The core of this definition lies in AC2. Informally, AC2(a) is inspired by the traditional counterfactual reasoning, but it is more permissive: it allows the dependence of φ on X to be tested under special *structural contingencies*, in which the variables $\vec{W}$ are held constant at some setting $\vec{w}$. AC2(b) ensures that setting $\vec{W}$ to $\vec{w}$ (or any subset of it) merely eliminates spurious side effects that tend to mask the action of X. AC1 just says that A cannot be a cause of B unless both A and B are true; and AC3 ensures minimality, preventing the situation, where, for example, rock-throwing and sneezing could be considered as a cause for the bottle shattering.

The requirement for AC2(b) to hold for all subsets of W prevents situations where W "conceals other causes for φ". The role of this requirement is perhaps best understood by considering the following example, due to Hopkins and Pearl [25] (the description is taken from [20]): Suppose that a prisoner dies either if A loads B's gun and B shoots, or if C loads and shoots his gun. Taking D to represent the prisoner's death and making the obvious assumptions about the meaning of the variables, we have that $D = (A \wedge B) \vee C$. Suppose that in the actual context u, A loads B's gun, B does not shoot, but C does load and shoot his gun, so that the prisoner dies. That is, $A = 1$, $B = 0$, and $C = 1$. Clearly $C = 1$ is a cause of $D = 1$. We would not want to say that $A = 1$ is a cause of $D = 1$, given that B did not shoot (i.e., given that $B = 0$). The key point is that AC2(b) says that for $A = 1$ to be a cause of $D = 1$, it must be the case that $D = 0$ if only some of the values in $\vec{W}$ are set to $\vec{w}$. That means that the other variables get the same value as they do in the actual context; in this case, by setting only A to 1 and leaving B unset, B takes on its original value of 0, in which case $D = 0$.

3.3 Responsibility and blame

The definitions of responsibility and blame below are by Chockler and Halpern [8].

DEFINITION 3.2. *The* degree of responsibility of $\vec{X} = \vec{x}$ for φ in $(M, \vec{u})$, *denoted* $\mathrm{dr}((M, \vec{u}), (\vec{X} = \vec{x}), \varphi)$, *is* 0 *if $\vec{X} = \vec{x}$ is not a cause of φ in $(M, \vec{u})$; it is $1/(k+1)$ if $\vec{X} = \vec{x}$ is a cause of φ in $(M, \vec{u})$ and there exists a partition $(\vec{Z}, \vec{W})$ and setting $(\vec{x}', \vec{w})$ for which AC2 holds such that (a) k variables in $\vec{W}$ have different values in $\vec{w}$ than they do in the context $\vec{u}$ and (b) there is no partition $(\vec{Z}', \vec{W}')$ and setting $(\vec{x}'', \vec{w}')$ satisfying AC2 such that only $k' < k$ variables have different values in $\vec{w}'$ than they do the context $\vec{u}$.*

Intuitively, *dr*$((M, \vec{u}), (\vec{X} = \vec{x}), \varphi)$ measures *the minimal number of changes* that have to be made in $\vec{u}$ in order to make φ counterfactually depend on $\vec{X}$, provided the conditions on the subsets of $\vec{W}$ and $\vec{Z}$ are satisfied. If there is no partition of $\mathcal{V}$ to $(\vec{Z}, \vec{W})$ that satisfies AC2, or $(\vec{X} = \vec{x})$ does not satisfy AC3 for φ in $(M, \vec{u})$, then the minimal number of changes in $\vec{u}$ in Definition 3.2 is taken to have cardinality ∞, and thus the degree of responsibility of $(\vec{X} = \vec{x})$ is 0 (and hence it is not a cause). Clearly, the larger the degree of responsibility is, the more influential is the cause on the value of φ.

EXAMPLE 3.3. *Consider the following example. Suppose there are 11 voters. Voter i is represented by a variable X_i, $i = 1, \ldots, 11$; the outcome is represented by the variable O, which is 1 if Mr. B wins and 0 if Mr. G wins. In the context where Mr. B wins 11–0, it is easy to check that each voter is a cause of the victory (that is $X_i = 1$ is a cause of $O = 1$, for $i = 1, \ldots, 11$). However, the degree of responsibility of $X_i = 1$ for is $O = 1$ is just $1/6$, since at*

least five other voters must change their votes before changing X_i to 0 results in $O = 0$. But now consider the context where Mr. B wins 6–5. Again, each voter who votes for Mr. B is a cause of him winning. However, now each of these voters have degree of responsibility 1. That is, if $X_i = 1$, changing X_i to 0 is already enough to make $O = 0$; no other variables need to change.

The definition of blame addresses the situation where there is uncertainty about the true situation or "how the world works". Blame, introduced in [8], considers the "true situation" to be determined by the context, and "how the world works" to be determined by the structural equations. An agent's uncertainty is modeled by a pair $(\mathcal{K}, \Pr)$, where $\mathcal{K}$ is a set of pairs of the form $(M, \vec{u})$, where M is a causal model and $\vec{u}$ is a context, and $\Pr$ is a probability distribution over $\mathcal{K}$. A pair $(M, \vec{u})$ is called a *situation*. We think of $\mathcal{K}$ as describing the situations that the agent considers possible before $\vec{X}$ is set to $\vec{x}$. The degree of blame that setting $\vec{X}$ to $\vec{x}$ has for φ is then the expected degree of responsibility of $\vec{X} = \vec{x}$ for φ in $(M_{\vec{X} \leftarrow \vec{x}}, \vec{u})$, taken over the situations $(M, \vec{u}) \in \mathcal{K}$. Note that the situation $(M_{\vec{X} \leftarrow \vec{x}}, \vec{u})$ for $(M, \vec{u}) \in \mathcal{K}$ are those that the agent considers possible after $\vec{X}$ is set to $\vec{x}$.

DEFINITION 3.4. *The* degree of blame of setting $\vec{X}$ to $\vec{x}$ for φ relative to epistemic state $(\mathcal{K}, \Pr)$, *denoted* $\mathrm{db}(\mathcal{K}, \Pr, \vec{X} \leftarrow \vec{x}, \varphi)$, *is*

$$\sum_{(M,\vec{u}) \in \mathcal{K}} \mathrm{dr}((M_{\vec{X} \leftarrow \vec{x}}, \vec{u}), \vec{X} = \vec{x}, \varphi) \Pr((M, \vec{u})).$$

3.4 Defaults and normality

In this section we review the problem that arises from concepts of defaults and normality and the ways to address it; the material is largely taken from [19]. While the definitions of causality, responsibility, and blame provide clear and quantified measures of causality and agree with our intuition in most cases, they only check the existence of a contingency that creates a counterfactual dependence; the "normality" of this contingency is never taken into account. This can lead to surprising results, as illustrated by the following example, taken from [24] (see also [23]).

EXAMPLE 3.5. *Assassin is in possession of a lethal poison, but has a last-minute change of heart and refrains from putting it in Victim's coffee. Bodyguard puts antidote in the coffee, which would have neutralized the poison had there been any. Victim drinks the coffee and survives.*

Out intuition says that Bodyguard's putting the antidote in the Victim's coffee is not a cause of the Victim being alive; however, according to Def. 3.1, it is a cause, for there is a (not very likely) contingency where Assasin attempts poisoning and the Victim survives iff there is an antidote in his coffee.

Ideally, we would like to only consider contingencies which change the values of variables to *more normal* values, according to our understanding of how the world works. [19] suggests to extend the causality framework by adding a *ranking function* that associates with each possible world a rank, reflecting its "normality". Then, the only contingencies we are allowed to consider are those that lower the current ranking of the world, that is, make it "more normal". This definition reflects our intuition by removing unlikely contingencies from consideration; under this definition, the antidote in the coffee is not a cause of the victim being alive, since not being poisoned is a more normal situation than being poisoned.

The ranking function can be combined with both the definition of responsibility and the definition of blame in a straightforward way. Since the responsibility is simply a quantitative measure of causality, if there exists a contingency in a more probable world that creates a causal dependence, the responsibility will measure the size of this contingency, as in the original definition. For blame, we can take into account only the *more normal* or *more probable* worlds, or, alternatively, we can also consider the higher ranked worlds, but attribute them a much lower probability.

4. CASE STUDY

Our main case study, on which we demonstrate our approach, is the "Baby P." case. We describe the case here, and construct the causal model for it in Section 5.

"Baby P", or Peter Connelly, was a baby who was born (March 2006), lived and died (August 2007) in the London borough of Haringey after suffering physical abuse over a sustained period of time. The circumstances leading up to baby Peter's death are particularly perplexing because (i) Baby Peter was listed in the Child Protection Register due to "physical abuse and neglect" and (ii) he and his carers were being actively monitored by a system of "joined-up governance" involving determinedly collaborating professionals and organisations who were all aware of (i) [31].

Baby Peter suffered the abuse while living in a home with three adults able to care for him: Peter's mother Tracey Connelly, Peter's stepfather Steven Barker and Steven Barker's brother Jason Owen. After Peter's death, all three denied abusing Peter or having witnessed any abuse. As a jury concluded, based on the available evidence, that at least one of the adults in the household injured Peter but was unable to identify which one, none were found guilty of murder or manslaughter, but all three were found guilty of "causing or allowing [Peter's] death" [32]. In other words, the Court decided that each of the three adults are equally responsible for the death of Peter and that each was in a position to have been able to prevent his death. It is worth noting that, in sentencing, the judge takes into account a number of factors, other than the degree of responsibility, such as mitigating circumstances, prior/related convictions and the extent to which each individual is a risk to the public.

Baby Peter was seen regularly by medical professionals, social workers and other professionals involved in his case. On a number of occasions, these professionals and their associated organisations had opportunities to remove Peter from the care of the people that caused or allowed the abuse to take place. Less than two days before his death, Peter was seen by a locum consultant[2] paediatrician at a "child development clinic", where Peter was referred as part of registration on the Child Protection Register. An autopsy performed after Peter's death revealed that Peter would have had a broken back and several other severe non-accidental injuries at this time. However, the doctor failed to perform a full examination because Peter was "miserable and cranky". Had she done so, the injuries would normally have been observed and Peter would have had to be admitted to hospital. A subsequent review by England's health and social care regulator, the Care Quality Commission, criticised the hospital where the doctor worked for "poor recruitment practices", "lack of specific training in child protection", "shortages of staff" and "failings in governance" [9], and this may have adversely affected the ability of the consultant paediatrician on duty to make sound decisions in this case.

On two occasions, once in December 2006 and once in June/July 2007, Peter was assessed at A&E[3]. On both occasions, (i) medical staff observed a range of injuries incurred at different times that appeared to be non-accidental, (ii) Peter was diagnosed not to

[2]A consultant working on a short term contract.

[3]The Accident and Emergency department of a UK hospital.

suffer a medical condition that would cause him to bruise easily, (iii) the medical examinations was followed by a police investigation, with input from social services, and (iv) Tracey Connelly was unable to provide a satisfactory explanation for what may have caused Peter's injuries. Following the December 2006 investigation, Haringey Legal Services (HLS) decide that the "threshold to initiate care proceedings", which could lead to a care order removing Peter from the care of his mother permanently, was met. In spite of this, Haringey's Children and Young People's Service decide not to proceed, but these proceedings did lead to Peter being listed on the Child Protection Register. Following the June/July 2007 investigation, HLS decided that the threshold to initiate care proceedings was not met. The police reported its findings to the Crown Prosecution Service, who found that there was insufficient evidence to pursue with a prosecution of Peter's carers.

The decisions made by Haringey Legal Services and the Crown Prosecution Service in determining whether to proceed with care proceedings or a prosecution of Peter's carers has been affected by the evidence provided by social workers and police, their review of the evidence and their policy/procedures to determine whether to proceed with an application for care order or a prosecution. Although he had been identified to police and social workers as an unnamed friend of Tracey Connelly, it was not until the investigation into Peter's death that Steven Barker's role as stepfather of Peter and cohabitant of Tracey Connelly was revealed. As a result, police investigators and social workers were confronted with conflicting evidence: on the one hand medical experts report that Peter must have sustained non-accidental injuries over a period of time, and on the other hand social workers observed that Peter appeared to have a good relationship with his mother who was, so police investigators and social workers believed at the time, his sole carer. But Tracey Connelly hid the identity and role of Steven Barker from police investigators and social workers and could have been challenged more robustly to explain the conflicting evidence [21]. Although Steven Barker had no prior convictions before the death of Peter, he was known to police to have a history of violence [14]. As such, one can argue that a more in-depth investigation by police, a greater willingness to question Tracey Connelly by social workers or a more robust review of the evidence by the legal services might have revealed information that could have lead to the start of legal proceedings.

5. MODELLING THE BABY P. CASE IN THE CAUSALITY FRAMEWORK

In this section we discuss ways to capture the Baby P. case as a causal model. This representation allows us to derive quantitative measures of responsibility and blame for involved parties. We start with the simplest model and gradually add more information to it, that allows us to compute the degree of responsibility and blame of the involved parties. The quantitative assessments that we compute here are, of course, just rough estimates for illustrative purposes. In fact, these details are exactly the type of thing that an inquiry could debate and perhaps gather evidence for. The process of gradual refinement of the model matches the process of collecting evidence, and we expect that this is the way the causal modelling will be used during the inquiry stage.

5.1 The simplest model – responsibility and blame of the adults in Baby P.'s household

Our first attempt is a simple causal model M_1 informally depicted in Fig. 2. This model focusses on the adults living in the same household with Baby P. The variables of this model are listed in Table 1.

Variable	Meaning	Value
PD	Baby Peter dies	1
TC	Tracey Connelly caused the death of Baby Peter	?
SC	Steven Barker caused the death of Baby Peter	?
JC	Jason Owen caused the death of Baby Peter	?
TA	Tracey Connelly allowed the death of Baby Peter	?
SA	Steven Barker allowed the death of Baby Peter	?
JA	Jason Owen allowed the death of Baby Peter	?
PMC	Baby Peter is allowed to be in the care of his mother at the time of his death	1

Table 1: Variables of the model M_1 and known values

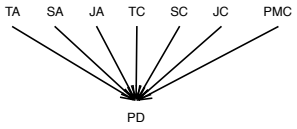

Figure 2: The simplest causal model M_1 for the Baby P. case

.

Based on the description of the case and our understanding of the world, the death of Baby P. could occur only if at least one of Tracey Connelly, Steven Barker and Jason Owen caused his death and all of them allowed it. In addition, it is required that Baby Peter is allowed to remain in the care of his mother (PMC). The following equation captures this dependency:

$$\begin{aligned} PD =& (TC \vee SC \vee JC) \wedge \\ & (TA \wedge SA \wedge JA) \wedge PMC. \end{aligned} \tag{1}$$

As the value of PMC is affected by factors outside of M_1, we consider it an *exogenous variable* at this stage, that is, a variable whose value is set externally. As external observers, our knowledge of the true situation is limited, as not all values of variables are known to us in advance; if they were, it would have been a much easier task for the court to determine the appropriate degree of responsibility of each of the involved parties. However, since $PD = 1$ and $PMC = 1$ (which we know independently from the value of PD), we can deduce the values of other variables with some degree of certainty. In particular, $PD = 1$ implies that

$$(TC \vee SC \vee JC) \wedge (TA \wedge SA \wedge JA) = 1, \tag{2}$$

hence

$$TC \vee SC \vee JC = 1 \tag{3}$$

and

$$TA \wedge SA \wedge JA = 1. \tag{4}$$

From Equation 3 follows that at least one of the variable TC, SC, JC has the value 1, and from Equation 4 follows that TA, SA, and JA have the value 1. We model our uncertainty as a pair $(\mathcal{K}_1, \Pr_1)$, as defined in Sec. 3, where $\mathcal{K}$, in our case, represents the possible contexts, describing situations (assignments of variables) for

Context	TC	SC	JC	$\Pr_2$
$\vec{u}_1^1$	1	0	0	0.1
$\vec{u}_2^1$	1	1	0	0.15
$\vec{u}_3^1$	0	1	0	0.4
$\vec{u}_4^1$	0	1	1	0.25
$\vec{u}_5^1$	0	0	1	0.05
$\vec{u}_6^1$	1	0	1	0
$\vec{u}_7^1$	1	1	1	0.05

Table 2: Situations and probabilities for M_1

M_1; since the model stays the same, we omit it from our notation. The function $\Pr_1$ is a probability distribution over the situations. There exist 7 possible assignments to the variables TC, SC, and JC that satisfy Equation 3 (this is because the total number of possible assignments is $2^3 = 8$, and only one of them – the one that assigns 0 to all three variables – does not satisfy Equation 3). Each such assignment represents a possible *situation* (see Sec. 3), or, in other words, our understanding about what happened. We can assign probabilities to these assignments based on our understanding of how the world works and the information gathered during the analysis of the case (for example, that it is less likely for a mother to cause death of her baby than for the mother's partner, and that Baby P. was attached to his mother, indicating that physical abuse from her side was not very likely). The probabilities capture the agent's (in this case, our) knowledge and understanding or could represent the evidence explicitly; we list these probabilities in Table 2; since TA, SA, and JA are assigned to 1 in all situations, we omit them from the table. While we attempt to match the probabilities with the details known about the case, they are still, of course, just rough estimates for illustrative purposes. In the future automated framework, we envision having a separate Bayesian network that represents the evidence and is used to compute posterior probabilities.

Note that the probabilities sum to 1, and that we assume that the probability of Tracey Connelly and Jason Owen causing the death of Baby P. is 0 (corresponding to $\vec{u}_6^1$), that is, we consider it not possible that both of them performed actions that caused the death of Baby P., while Steven Barker did not. We also assign very low probabilities to the situations $\vec{u}_5^1$ and $\vec{u}_7^1$. The situation $\vec{u}_5^1$ captures the possibility that Jason Owen alone caused the death of Baby P.; we consider this unlikely, because the first injuries of Baby P. were reported before Jason Owen moved into the house. The situation $\vec{u}_7^1$ captures the possibility that all three adults living in the house performed actions that caused the death of Baby P.; we consider it unlikely from our general understanding of how the world works.

The degree of responsibility *dr* of each of the variables in the value of PD is computed using Def. 3.1. We note that there is a counterfactual dependence between the value of TC and the value of PD in $\vec{u}_1^1$, corresponding to the situation, where Tracey Connelly alone caused the death of Baby P., and similarly for the value of SC in $\vec{u}_3^1$ and the value of JC in $\vec{u}_5^1$. Hence, in these contexts, the responsibility of the corresponding variable is 1. In other contexts, change in one or more values of other variables is required in order to create a contingency where such a counterfactual dependence exists (essentially, a change creating the contingency should satisfy the condition AC2 in Def. 3.1). In contexts, in which two variables are assigned 1, the size of the minimal change is 1, as changing the value of one variable to 0 creates a counter-factual dependence between the value of PD and the value of the other variable. Hence the responsibility of each of the variables assigned 1 in these contexts is $1/2$. Now consider the context $\vec{u}_7^1$, where we assume that all three adults living in the house – Tracey Connelly, Steven Barker, and Jason Owen – caused the death of Baby P. In this case, we need to change the values of two variables in order to create the counterfactual dependence, hence the responsibility of each variable in $\vec{u}_7^1$ is $1/3$.

The *degree of blame* of each of the three adults living in the house is the expected degree of responsibility computed according to Def. 3.4 as follows, where $(\mathcal{K}_1, \Pr_1)$ captures our epistemic state with respect to the model M_1:

$$\begin{aligned} db(\mathcal{K}_1, \Pr_1, TC \leftarrow 1, PD) =& 1 \times 0.1 + 1/2 \times (0.15 + 0) + \\ &+ 1/3 \times 0.05 \approx 0.19; \\ db(\mathcal{K}_1, \Pr_1, SC \leftarrow 1, PD) =& 1 \times 0.4 + 1/2 \times (0.15 + 0.25) + \\ &+ 1/3 \times 0.05 \approx 0.62; \\ db(\mathcal{K}_1, \Pr_1, JC \leftarrow 1, PD) =& 1 \times 0.15 + 1/2 \times (0.1 + 0) + \\ &+ 1/3 \times 0.05 \approx 0.19. \end{aligned} \tag{5}$$

In other words, we deduce that all three adults living in the household have a non-zero blame in causing the death of Baby P., and that Steven Barker's blame is higher than Tracey Connelly's and Jason Owen's (note that both responsibility and blame can range from 0 to 1, where 0 means not responsible or not to blame).

Now consider the responsibility of Tracey Connelly, Steven Barker, and Jason Owen in allowing Baby P.'s death. Since the value of PD counterfactually depends on Equation 4, we conclude that there is a counterfactual dependence between the value of PD and the values of each of the variables TA, SA, and JA. Hence, the only possible values of these variables are 1, and therefore the degree of blame coincides with the degree of responsibility, which is 1 for each of these variables.

To summarize our analysis of the blame of Tracey Connelly, Steven Barker, and Jason Owen, we conclude that Tracey Connelly's and Jason Owen's blame in causing Baby P.'s death is 0.19 each, Steven Barker's blame is 0.57, and that all three of them are 1-responsible and have a degree of blame 1 (meaning fully responsible and having full blame) for allowing Baby P.'s death. Note that neither the degrees of responsibility nor the degrees of blame are expected to sum to 1, in the general case.

5.2 A more detailed model – responsibility and blame of Dr Al-Zayyat

The next step in refining the model is introducing some additional variables that affect the value of PMC, and in particular, the role of Dr Al-Zayyat. This more detailed model, which we denote M_2, is depicted in Fig. 3. The additional variables that appear in this model are listed in Table 3. The dependencies between the variables in the model M_2 and the equations capturing these dependencies are as follows:

1. Baby Peter is left in the care of his mother if he is not admitted to the hospital by the doctor on duty at the Child Development Clinic where Baby Peter was seen, the court does not issue a care order (following the Dec 2006 events), and Haringey's Children & Young People Service does not apply for a care order at the second opportunity (following the Jul 2007 events). Formally, $PMC = \neg DAH \wedge \neg CCO1 \wedge \neg HCA2$.

2. The doctor on duty at the Child Development Clinic where Baby Peter was seen would have admitted Baby Peter to hospital if Dr Al-Zayyat had performed a full investigation of Baby Peter, or had the doctor on duty been sufficiently experienced and trained to run Child Development Clinics (in

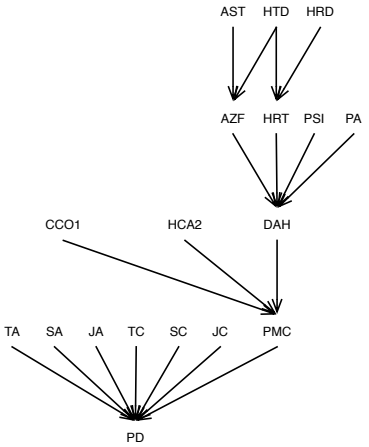

Figure 3: A more refined model M_2 for the Baby P. case

.

which case, a full-examination would have been performed in the case of Baby Peter), and, in addition, either Baby P. would have found to be severely injured or there was sufficient evidence to him being abused. Formally, $DAH = (AZF \vee HRT) \wedge (PSI \vee PA)$.

3. HRD refers to the experience of the doctor that is recruited by the hospital to run the Child Development Clinics, and has the value 0, indicating that an inexperienced doctor was recruited, based on the evidence in the case regarding the level of experience of Dr Al-Zayyat. HTD refers to the adequacy of the training provided. That is, $HRT = HTD \vee HRD$.

4. Dr Al-Zayyat might have performed a full-examination of Baby Peter, had she taken a sufficiently thorough approach to examinations in the clinic, or if she had received adequate training, which would have instilled in her the need to examine babies such as Baby Peter fully, irrespective of their crankiness. Formally, $AZF = (AST \vee HTD)$.

Since there is a counterfactual dependence between the value of PD and the value of PMC (see Equation 5.1), and between the value of PMC and the value of DAH in order to compute Dr Al-Zayyat's degree of responsibility and blame in the death of Baby P., it is enough to analyse the equations that define the value of DAH (Baby P. is admitted to the hospital).

While we know that, in fact, Baby P. was already severely injured at the time of his visit to the child's development clinic, and that his injuries indicate that he was abused, the materials of the case suggest that Dr Al-Zayyat was unaware of this fact. The materials of the case suggest, however, that there was enough data to at least put Baby P. at the risk of abuse and to ensure he is examined thoroughly. In other words, while Dr Al-Zayyat did not know the values of PSI and PA in advance, she should have attributed to them high enough probability to insist on thorough examination.

In this part of the model, *the uncertainty reflects the knowledge and beliefs of Dr Al-Zayyat*. We model her uncertainty as a pair $(\mathcal{K}_2, \Pr_2)$, where $\mathcal{K}$ represents the possible contexts for M_2, and the function $\Pr_2$ is a probability distribution over the situations, based on what Dr Al-Zayyat *should have known*. Based on the analysis above, the equation for DAH that incorporates all the independent variables is

$$DAH = (AST \vee HTD \vee HRD) \wedge (PSI \vee PA). \quad (6)$$

Variable	Meaning	Value
DAH	Child development clinic doctor admits Baby Peter to hospital to treat his injuries	0
AZF	Dr Al-Zayyat performs a full physical examination of Baby Peter on 1/8/2007	0
PSI	Baby Peter is severely injured on 1/8/2007	1
PA	Baby Peter is abused at home	1
HRT	The hospital recruited a doctor that was sufficiently experienced to run the child development clinic and provided adequate training	0
AST	Dr Al-Zayyat is sufficiently thorough in examining children at the child development clinic	0
HRD	Experienced doctor recruited for the child development clinic	0
HTD	Adequate training is provided to the doctor recruited for the child development clinic	0
$CCO1$	Court makes a care order	0
$HCA2$	Haringey's Children & Young People Service applies for a care order at the second opportunity in July 2007	0

Table 3: Additional variables of M_2 and known values

Context	PSI	PA	$\Pr_2$
$\vec{u}_1^2$	1	1	0.2
$\vec{u}_2^2$	0	1	0.1
$\vec{u}_3^2$	1	0	0.1
$\vec{u}_4^2$	0	0	0.6

Table 4: The epistemic state of Dr Al-Zayyat for M_2

Given that the values of HTD and HRD are 0, and the values of PSI and PA are 1, there is a clear counterfactual dependence between the value of AST, that represents the thoroughness of Dr Al-Zayyat, and the value of DAH, that represents the decision to admit Baby P. to the hospital (which was not taken). Therefore, based on the complete knowledge of the situation, Dr Al-Zayyat is 1-responsible (*fully responsible*) for not admitting Baby P. to the hospital, and hence for his death. However, when computing the blame of Dr Al-Zayyat, we consider several possible situations with respect to the values of PSI and PA, as listed in Table 4. The probabilities assigned to these situations reflect what Dr Al-Zayyat *should have known* (with the caveat that the numbers are, again, for illustrative purposes only). Since AST, HTD, and HRD all have the value 0, we omit them from the table.

The degree of responsibility *dr* of the value of AST in the value of DAH is computed using Def. 3.1. In situations $\vec{u}_1^2$, $\vec{u}_2^2$, and $\vec{u}_3^2$, there is a counterfactual dependence between the value of AST and the value of DAH, based on Equation 5.2, hence the degree of responsibility is 1 in these situations. In $\vec{u}_4^2$, the minimal change that satisfies the condition AC2 in Def. 3.1 is 1 — changing the value of either PSI or PA to 1 — hence the degree of responsibility of the value of AST in the value of DAH in $\vec{u}_4^2$ is 0.5.

Similarly to the computation in Sec. 5.1, the *degree of blame* of the value of AST in the value of DAH is the expected degree of responsibility computed according to Def. 3.4 as follows, where $(\mathcal{K}_2, \Pr_2)$ captures the epistemic state of Dr Al-Zayyat with respect

Context	HTD	HRD	PSI	PA	$\Pr_2'$
$\vec{u}_1^2$	0	0	1	1	0.1
$\vec{u}_2^2$	0	0	0	1	0.2
$\vec{u}_3^2$	0	0	1	0	0.1
$\vec{u}_4^2$	0	0	0	0	0.3
$\vec{u}_5^2$	1	1	0	0	0.1
$\vec{u}_6^2$	0	1	0	0	0.1
$\vec{u}_7^2$	1	0	0	0	0.1

Table 5: The extended epistemic state for M_2

to the model M_2:

$$db(\mathcal{K}_2, \Pr_2, AST \leftarrow 0, \neg DAH) = \\ =1 \times (0.2 + 0.1 + 0.1) + 1/2 \times 0.6 = 0.7. \quad (7)$$

In other words, we deduce that Dr Al-Zayyat's blame in not admitting Baby P. to the hospital, a decision that ultimately led to Baby P.'s death, is 0.7.

This computation does not take into account the values of HTD and HRD, mirrorring the process of public inquiry, where details of the case are discovered gradually. We now show that the model can be extended to consider the values of HTD and HRD as well. Considering all possible situations with respect to the values of HTD, HRD, PSI, and PA would have resulted in a table with $2^4 = 16$ rows. However, the number of situations in which DAH has the value 0 is significantly lower, as in order for DAH to have the value 0, either $(AST \vee HTD \vee HRD)$ or $(PSI \vee PA)$ should be 0, resulting in 7 situations in total, as depicted in Table 5 (again, the probabilities are for illustrative purposes only).

The degree of responsibility of the value of AST in the value of DAH in $\vec{u}_5^2, \vec{u}_6^2$, and $\vec{u}_7^2$ is 0, as there is no contingency that creates a counterfactual dependence between the value of AST and the value of DAH, while still satisfying the condition AC2(b) of Def. 3.1. The degree of responsibility of of the value of AST in the value of DAH in situations $\vec{u}_1^2, \vec{u}_2^2, \vec{u}_3^2$ and $\vec{u}_4^2$ is the same as before. Hence, taking the updated probabilities $\Pr_2'$ into account, it is easy to see that the updated degree of blame of Dr Al-Zayyat is

$$db(\mathcal{K}_2, \Pr_2', AST \leftarrow 0, \neg DAH) = \\ =1 \times (0.1 + 0.2 + 0.1) + 1/2 \times 0.3 = 0.55. \quad (8)$$

Note that the degree of responsibility stays the same, as it relies on the values of the variables in the real world.

REMARK 5.1. *The decisions of Dr Al-Zayyat can be viewed as an illustration on the concept of* defaults and normality *discussed in Sec. 3.4. The case indicates that she assumed that Baby P. was neither seriously injured, nor abused at the time he was seen in the clinic. While there exists a contingency that satisfies the condition AC2 of Def. 3.1, Dr Al-Zayyat might have* ranked *this contingency higher (in other words, being* less normal*) than the perceived situation, where Baby P. was assumed to be suffering of a minor illness. For legal purposes, however, we take into account not what the agent knows, but what she* should have known*, and in this case, the expectation is that Dr Al-Zayyat should have considered a possibility of a situation in which Baby P. was indeed seriously injured or abused.*

5.3 Further extensions of the model – other involved parties and timeline

The model can be further extended by considering the actions of other involved parties, notably, Haringey's Children & Young People Service, Haringey Legal Services, the police, and the court. The variable $CCO1$ depends on the actions of involved parties, including the judge, independent experts, the police, the result of the investigation of Haringey Legal Services, and the recommendation of social services. Similarly, the recommendation of Haringey's Children & Young People Service depends on the thoroughness of their examination of Baby P.'s situation; the fact that Steven Barker was living in the house was discovered only after Baby P.'s death, indicating a possible lack of thoroughness on their side.

An additional dimension is the timeline, or, more accurately, the order of events with respect to baby Peter's well-being. The order of events can be captured by introducing new, auxiliary variables that express this order, similarly to the detailed model of the example in Section 3 (see Fig. 1 on the right). In that example, the variables SH (Suzy hits) and BH (Billy hits) captured the order between the events of both rock throws hitting the bottle. In the case of Baby P., we can introduce additional variables $PMMYY$, where MM stands for the month, and YY for the year, and the range of the variables is { Thriving, Neglected, Abused}.

The equation for PD will then incorporate the variables $PMMYY$, based on our understanding that baby Peter's death resulted from abuse, and that there was no single incident of abuse, but rather a prolonged series of abuse and neglect over the course of several months, culminating in baby Peter's death. In this, more accurate model, it is possible to introduce the equations capturing the dependence of the values of $PMMYY$ on the behaviour of the adults living with baby Peter as well as the official role holders, such as the GP, social services, the police, and the court. Due to the lack of space and lack of detailed information about baby P.'s condition at different times, we defer this model to future work.

Finally, we note that every subsequent refinement we introduce in the model does not affect the previously computed degrees of responsibility and blame. This is due to the fact that the model is highly *decomposable*: it considers several very different aspects of blame in Baby P.'s death. By considering each aspect separately, we do not invalidate the previously computed results.

6. TOWARDS A DECISION SUPPORT TOOL AND EXTENSIONS OF THE ANALYSIS

The process of constructing the causal model and the equations involves an in-depth analysis of the facts of a case, similar to that performed in a public inquiry. However, the findings of the analysis need to be represented in the form of a rigorous model. While this complicates the analysis somewhat, it also leads to a clearer and unambiguous specification of the findings, so that these can be scrutinised subsequently.

Computing the degrees of responsibility and blame can be automated based on the suggestions we made in this paper. Indeed, given a model, a set of situations, and the accompanying probabilities, the degrees of responsibility and blame can be computed directly using the definitions in Sec. 3. However, since the complexity of the exact computation is quite high (see [5]), a straightforward implementation is likely to lead to prohibitively heavy computations. While better algorithms remain to be developed, we expect them to use the following observations to improve their efficiency. First of all, models are expected to be quite small by computational standards, even for very complex cases, hence even a "brute-force" analysis (checking all possible contingencies one after another) may finish in a reasonable amount of time. Second, the dependencies between variables are far from random, as they depict the dependencies between people's actions and outcomes in the real world. As we saw in Sec. 5, and we predict that we will see this situation fairly often, cases can be decomposed in a straight-

forward way, by considering every conjunct of a large causal formula separately, thus making the computation much faster. In legal reasoning this decomposition corresponds to considering different parties in the case separately, mirroring the process of inquest and inquiry. Furthermore, we expect the models constructed from real legal cases to be amenable to heuristics similar to those used in SAT solvers [17], which have been shown to dramatically decrease the complexity of solving instances of the satisfiability problem constructed from real industrial cases.

We propose to build an automated framework for computing the degrees of responsibility and blame, with a user-friendly graphic user interface (GUI), that will allow the users to input the data employing an appropriate modelling tool. The causal models will be constructed automatically by the framework, taking into account the dependencies between the variables, known values, and the knowledge or belief of the user about the unknown values. The framework can then compute the probability of each situation based on the probabilities of the values of variables, and compute the degree of responsibility and blame of the involved parties based on this data.

In real-world use of our approach, there is likely to be substantial uncertainty or disagreement between stakeholders concerning the subjective probability assessment the approach relies on. Indeed, relying solely on point probabilities is likely to constitute a barrier to the adoption of this approach. However, it is important to note that the framework proposed herein does not require the use of point probabilities, and critical conclusions can still be drawn with more qualitative or ordinal assessments [26], or working with thresholds. The formal framework also allows for sensitivity analyses [4], where the impact of different assumptions or estimates can be assessed. Alternatively, the approach can be extended by modelling an agent's uncertainty as $(\mathcal{K}, \mathrm{Kr})$, where Kr is a *credal* set: i.e. a convex set of probability distributions over situations [10]. Recent work has shown how argumentation models can be constructed to justify constraints over probability distributions [27]. In the latter approach, sets of constraints justified by sets of accepted arguments entail credal sets, thereby enabling the convex sets of subjective probability distributions employed in models to be scrutinised by means of computational models of argumentation.

The framework supports output as quantification of responsibility and blame, which is useful in complex civil cases, where there is a need to determine the amounts paid to the claimant by each of the involved parties (see Sec. 2). In other cases, it might be more useful to compute, whether the degree of responsibility (or the degree of blame) is above a certain threshold, which can be considered as a threshold for requiring a response. For example, the court may decide that only parties with the degree of blame above 0.1 shall be prosecuted. This might be especially relevant in cases involving several further removed levels of responsibility, as described in Sec. 2. An additional attraction of using thresholds instead of exact values is that the computational effort involved in determining whether the degree of responsibility or blame is above a predefined threshold is much lower (linear in the size of the model) than the effort required to compute the exact degree of responsibility or blame.

7. CONCLUSIONS AND ROADMAP

Using a complex legal case we have shown how the proposed framework can capture complex causal interactions between multiple parties and events, and accommodate probabilistic information and assumptions. Given these assumptions, the model produces graded values of responsibility and blame for each party, detailing how much their actions (or omissions) were responsible for the target event (in this case the death of Baby P). Using illustrative probability values, the model assigns each defendant a different degree of blame for causing Baby P's death (a relative ordering that accords with the legal ruling), whereas all three defendants are fully to blame for allowing the death. The model also assigns degrees of blame to the numerous different parties who were involved in the case. One feature of the model is that it does not treat responsibility or blame as a fixed amount that is distributed amongst each agent; rather, multiple agents can all have full responsibility or blame for the same event, as is the case for the three defendants' responsibility (and blame) for allowing Baby P. to die. Clearly, this is just a roadmap, describing a promising direction that can help in structuring the considerations of causes, responsibility and blame in legal settings.

As mentioned in the introduction, we envisage that this framework, and the detailed causal analysis it permits, is most applicable at the attributive stage of a legal inquiry; for example, when considering sentencing or compensation, especially when multiple parties are involved, or when a broader inquiry is required, where a wider range of individuals or organisations might be held accountable even if not legally charged. Examples of the latter include inquiries into child abuse cases, medical negligence, police misconduct, phone-hacking and large-scale riots. In such cases, it is important to note that the graded assessments of blame and responsibility ought to be considered in conjunction with an assessment of the other factors that can affect *legal* responsibility, which is a broader concept than causal responsibility and blame [30]. In particular, the analysis of causality illustrated here on the Baby P. case is inherently *jurisdiction-dependent*, as the knowledge of actors or a their psychological of actors might be relevant in some jurisdictions and irrelevant in others. Fortunately, the proposed framework is flexible enough to accommodate different settings.

The legal profession is conservative when it comes to adopting new ideas and technology; the recent experience of the use of Bayes' rule in the law [6] is especially pertinent when it comes to the likely challenges to be faced in getting the ideas presented in this paper accepted. The arguments both in favour of and against the use of Bayes' rule apply equally to the proposed formal method of assigning responsibility. The argument for Bayes' rule is that, given some prior and conditional probabilities for hypotheses and related pieces of evidence (whose values are genuinely open to interpretation), the posterior probabilities can *only* be rationally computed by Bayes' rule. In arguing against this many legal professionals wrongly assume that the probability "updating" – like the prior probabilities – should be a matter for individual jurors/judges and there should be no reliance on a "mathematical formula". The method we have proposed for assigning responsibility similarly seeks to "mathematise" part of the legal process that lawyers (wrongly, in our opinion) feel is the sole domain of a judge's intuition.

Acknowledgements

The authors thank Joe Halpern for helpful discussions. The second author gratefully acknowledges the support of the ERC-2013-AdG339182-BAYES-KNOWLEDGE project.

8. REFERENCES

[1] Fairchild v Glenhaven Funeral Services Ltd. UKHL 22, 2002.

[2] Department of Health and The Home Office, The Victoria Climbié inquiry, Report of an Inquiry by Lord Laming, Cm 5730, January 2003.

[3] The Shipman inquiry. Third report. Command paper, Cm 5854, 2003.

[4] C. Aitken and F. Taroni. *Statistics and the Evaluation of Evidence for Forensic Science*. Wiley, 2nd ed. edition, 2004.

[5] G. Aleksandrowicz, H. Chockler, J. Y. Halpern, and A. Ivrii. The computational complexity of structure-based causality. In *Proceedings of the 28th AAAI*, pages 974–980, 2014.

[6] C. E. H. Berger, J. Buckleton, C. Champod, I. Evett, and G. Jackson. Evidence evaluation: A response to the court of appeal judgement. *R v T. Science and Justice*, 51:43–49, 2011.

[7] F. J. Bex, P. J. van Koppen, H. Prakken, and B. Verheij. A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law*, 18(2):123–152, 2010.

[8] H. Chockler and J. Halpern. Responsibility and blame: a structural-model approach. *JAIR*, 22:93–115, 2004.

[9] C. Q. Commission. Review of the involvement and action taken by health bodies in relation to the case of Baby P., May 2009.

[10] F. Cozman. Credal networks. *Artificial Intelligence*, 120:199–233, 2000.

[11] N. E. Fenton, D. Berger, D. A. Lagnado, M. Neil, and A. Hsu. When 'neutral' evidence still has probative value (with implications from the Barry George Case). *Science and Justice*, 54(4):274–287, 2014.

[12] N. E. Fenton, D. A. Lagnado, and M. Neil. A general structure for legal arguments using bayesian networks. *Cognitive Science*, 37:67–102, 2013.

[13] R. Fumerton and K. Kress. Causation and the law: Preemption, lawful sufficiency, and causal sufficiency. *Law and Contemporary Problems*, 64:83–105, 2001.

[14] C. Gammell. Baby P: Steven Barker and Jason Owen, brothers with a history of violence, 2009.

[15] T. Gerstenberg and D. Lagnado. Spreading the blame: the allocation of responsibility amongst multiple agents. *Cognition*, 115:166–171, 2010.

[16] R. Goldberg. *Perspectives on causation*. Hart publishing, 2011.

[17] C. P. Gomes, H. Kautz, A. Sabharwal, and B. Selman. Satisfiability solvers. In F. van Harmelen, V. Lifschitz, and B. Porter, editors, *Handbook of Knowledge Representation*, pages 89–123. Oxford University Press, Elsevier, 2008.

[18] N. Hall. Two concepts of causation. In J. Collins, N. Hall, and L. A. Paul, editors, *Causation and Counterfactuals*. MIT Press, Cambridge, Mass., 2002.

[19] J. Y. Halpern. Defaults and normality in causal structures. In *Proc. 11th KR*, pages 198–208. AAAI Press, 2008.

[20] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach. Part I: Causes. *British Journal for Philosophy of Science*, 56(4):843–887, 2005.

[21] L. S. C. B. Haringey. Serious Case Review: Baby P., 2009.

[22] H. L. A. Hart and T. Honoré. *Causation in the Law*. Oxford University Press, Oxford, U.K., second edition, 1985.

[23] E. Hiddleston. Causal powers. *British Journal for Philosophy of Science*, 56:27–59, 2005.

[24] C. Hitchcock. What's wrong with neuron diagrams? In J. Campbell, M. O'Rourke, and H. Silverstein, editors, *Causation and Explanation*, pages 69–92. MIT Press, Cambridge, MA, 2007.

[25] M. Hopkins and J. Pearl. Clarifying the usage of structural models for commonsense causal reasoning. In *Proc. AAAI Commonsense Symp.*, 2003.

[26] J. Keppens. Towards qualitative approaches to bayesian evidential reasoning. In *In Proc. of the 11th ICAIL*, pages 17–25, 2007.

[27] J. Keppens. On modelling non-probabilistic uncertainty in the likelihood ratio approach to evidential reasoning. *Artificial Intelligence and Law*, 22(3):239–290, 2014.

[28] D. Lagnado, T. Gerstenberg, and R. Zultan. Causal responsibility and counterfactuals. *Cognitive Science*, 37:1036–1073, 2013.

[29] D. A. Lagnado, N. E. Fenton, and M. Neil. Legal idioms: a framework for evidential reasoning. *Argument and Computation*, 4(1):46–63, 2013.

[30] J. Lehmann, J. Breuker, and J. Brouwer. Causation in AI and Law. *Artificial Intelligence and Law*, 12:279–315, 2004.

[31] M. Marinetto. A Lipskian analysis of child protection failures from Victoria Climbié to 'Baby P': A street-level re-evaluation of joined-up governance. *Public Administration*, 89(3):1164–1181, 2011.

[32] S. Remarks. The queen -v- (b) (the boyfriend of Baby Peter's mother) (c) (Baby Peter's mother) and Jason Owen, 2009.

[33] J. Stapleton. Choosing what we mean by 'causation' in the law. *Missouri Law Review*, 73:433–480, 2008.

[34] F. Taroni, C. Aitken, P. Garbolino, and A. Biedermann. *Bayesian Networks and Probabilistic Inference in Forensic Science*. John Wiley & Sons, Chichester, UK, 2nd edition, 2014.

[35] R. Wright. Causation in tort law. *California Law Review*, 73:1737–1828, 1985.

[36] R. Zultan, T. Gerstenberg, and D. Lagnado. Finding fault: causality and counterfactuals in group attributions. *Cognition*, 125(3):429–440, 2012.