

Evidential Reasoning about Motives: a Case Study

Floris Bex* Trevor Bench-Capon† Katie Atkinson‡

Abstract

Motives play an important role at every stage of a criminal investigation. They can be used to search for an explanation of the crime (why was this person killed?), to identify a suspect (who would have killed this person for this reason), and to persuade a jury of a suspect's guilt (this motive explains why this person committed this crime). We have previously developed an account of motivations based on a general approach to practical reasoning. In this paper we will concentrate on the use of motives to provide plausibility to a story intended to persuade a jury of a person's guilt or innocence. We will concentrate on a particular case study, formalised previously by Thagard, the two trials of Claus von Bülow. An advantage of our approach is that it allows stories to be considered from an *intentional* as well as a *physical* stance, whereas in previous accounts, including Thagard's, only the physical stance is available. We show how our approach can be used to explain the outcome of both trials, and to identify the points in which the defence could be improved from the first trial to the second.

1 Introduction

Motives play an important role at every stage of a criminal case. They can be used to search for an explanation of the crime (why was this person killed?), to identify a suspect (who would have the motive to kill this person) or to persuade a jury of a suspect's guilt (this motive explains why the suspect committed the crime). In this paper we present a formal framework for the analysis of reasoning with such motives and their underlying values, concentrating on the use of motives to provide plausibility to a story intended to persuade a jury of a person's guilt or innocence.

A persuasive story should be plausible in that it conforms to our beliefs about how things generally happen in the world around us. This plausibility is

*University of Groningen, Faculty of Law, The Netherlands. Email: f.j.bex@rug.nl

†Department of Computer Science, The University of Liverpool. Email: tbc@liverpool.ac.uk

‡Department of Computer Science, The University of Liverpool. Email: katie@liverpool.ac.uk

partly dependent on the plausibility of the (physical) causal relations between the events in the story. For example, a story in which the victim ends up in a coma because she was injected with insulin is only plausible if we believe that an overdose of insulin can cause such a comatose state. The plausibility of a story is also dependent on how likely it is that the agents in the story would have made the alleged choice in the situation, and this in turn depends on the value preferences we believe an agent to have. For example, a story in which the suspect kills the victim in order to inherit the victim's money is only plausible if we believe that the suspect values money higher than the life of the victim. When determining the plausibility of a story we can thus differentiate between plausibility as regarded from the *physical* as well as the *intentional stance* [8].

Other work on evidential reasoning in AI and Law, most notably by Bex and colleagues [6] and Thagard [11], mainly focuses on reasoning from the physical stance. While both approaches allow for the inclusion of links that denote some sort of "motivational causation" (i.e. motive *causes* action), the reasoning about why a certain choice was made by the agents in a case remains implicit. Similarly, Walton and Schafer [13] establish the existence of a motive, but lack the machinery to explain why an actor chose to act on this motive in the particular situation.

Our current account of reasoning with motives and values, which is based on a general approach to practical reasoning, was previously introduced in [4]. In this work we discussed how explanations of what happened in a criminal case can be inferred through abductive reasoning, and how the possible motives of the agents can influence the choice between these explanations. In this paper we will forgo further discussion of abductive inference. Instead, we will mainly look at how motives and agents' value preferences can be used to persuasively argue for one particular story in a case where there are multiple accounts of the events.

In the current paper we will also provide a study of an actual legal case, namely the two trials of Claus von Bülow, who was accused of attempting to murder his wife Sunny.¹ The case has been the subject of much discussion and even a Hollywood movie. It was previously formalised in AI and Law by Thagard [11], which allows us to actively compare our approach to Thagard's explanatory coherence theory.

In section 2 we will discuss the typical aspects of motives and values in a criminal legal context as far as relevant for this paper and in section 3 we recapitulate our approach to reasoning about motives. Section 4 will present the case study, and discuss the first trial in particular. Section 4 will show how our analysis can help to explain the different presentation of the defence case in the second trial, and why that presentation was successful. Section 5 will provide some discussion and concluding remarks.

¹Although the alleged crime took place in 1980, the case attracted some renewed interest in late 2008 when the victim died after 28 years in a coma.

2 Motives in evidential reasoning

When discussing reasoning about motives, it is useful to distinguish the separate concepts of *value*, *motive* and *goal*. Values, as defined by [3], can be seen as abstract principles which an agent or a group of agents hold. Examples of values are “wealth”, “love” and “honesty”. Agents can be expected to, actively or passively, promote the values which they find important. Our concept of motive is roughly the same as this concept of value, in that a motive is an abstract good which an agent may or may not want to promote. Such a motive can cause a person to form any number of goals in order to promote the principle for which the motive stands. For example, if Claus is motivated by his will to be independent from Sunny, he can plan to murder Sunny. However, he can also form the less drastic plan to simply divorce his wife. So the same motive can cause a person to form different goals. This current notion of motive and goal is somewhat different from [13], where the authors essentially equate an agent’s motive with an immediate goal of the agent. In our opinion, our current notion of motive better captures the basic intuition that the same motive can be satisfied in different ways. In the rest of this paper, the terms *motive* and *value* will often be used interchangeably.

Motives and values play an important role in all stages of criminal investigation and decision making. De Poot *et al.* [7] divide cases into several categories, where each category corresponds to a particular phase in the process of criminal investigation and decision making. In a *search case*, there is no suspect and no scenario. An example is a case in which only physical evidence is found, for example a dead body. In such a case, possible scenarios should be constructed and compared; in the example, scenarios for accident, suicide and murder can be constructed. An alleged motive can play an important part in constructing these provisional scenarios. If, for example, the person that is found dead is a wealthy heiress, one could do well by searching for scenarios that contain monetary motives on the part of the culprit. In an *investigation case*, the scenario is known but the suspect is not and it has to be investigated who could have been the main actor (i.e. the perpetrator) in the story about the crime. For example, this could involve looking for persons who would financially benefit from the heiress’ death, that is, persons for whom the motivation of money is a real option. In a *verification case*, the scenario as well as the identity of the possible suspect is known and this identity needs to be verified by providing a suitable motive. For example, the victim’s husband could be the kind of person who is motivated by monetary gain to kill his wife.

In all stages it is important to consider possible alternatives so as to avoid the well-known problem of confirmation bias or tunnel vision. One way of providing such an alternative is to tell a totally new story (with possibly a different suspect) that explains the evidence at least as well as the current story. Another way of providing an alternative is to argue that this particular suspect was not the perpetrator of the crime. Instead of providing a totally new explanation for the evidence, it is simply argued that the current suspect could not or would not commit such a crime, thus providing what we call a

“suspect-specific” alternative. For example, it can be argued that it was not physically possible for the suspect to have committed the crime because he or she was somewhere else at the time, thus providing the suspect with an alibi. It can also be argued that the suspect is not the type of person who would commit such a crime by showing that his motivational preferences are different from those assumed; in the example, one could argue that the husband values the love for his wife higher than money so he would never kill her, even though he would benefit financially from her death. Finally, it can be argued that given the suspect’s values, he could have satisfied his motives by acting differently. For example, it might be argued that it would have been easier for the suspect to divorce his wife and take half of her fortune with him.

These last two ways of providing a suspect-specific alternative are examples of reasoning about the suspect’s motives and the suspect’s willingness to act on such motives. This shows that in any (formal) reasoning framework which is used in evidential reasoning we should be able to rationalise why the suspect in a case chose to act on his motives in that particular situation.

Before we continue, one related notion, namely that of *character*, should be briefly discussed. In this paper, a person’s character is modelled as an ordering of his or her motives. For example, we can say that Claus is the kind of person who prefers independence over honesty. If we have evidence that Claus is such a person - for example, that he has lied on previous occasions - we can provide an argument for this particular preference. As Walton and Schafer [13] note, character evidence is often inadmissible at trial. However, it may be admissible for the purpose of proving someones motive, which is exactly the purpose we use it for.

3 Practical Reasoning with Motives

In [4], which has been extended and revised as [5], we attempted to analyse motives in terms of an agent choosing to act so as to promote some value by realising some goal, an approach to practical reasoning previously developed by Atkinson and her colleagues in [1], and [2]. The approach is based on the use of an argumentation scheme and critical questions which allow justifications for action to be presumptively proposed and critiqued. The original scheme defined by Atkinson et al. in [2], was as follows:

AS1 In the current circumstances R
 We should perform action A
 Which will result in new circumstances S
 Which will realise goal G
 Which will promote value V.

This argumentation scheme can be used by an agent to justify an action in particular circumstances in terms of his values. For a given instantiation of this scheme a number of critical questions could be posed, in order to challenge the various elements of the scheme and so dispute the presumptive conclusion. The sixteen listed in [2] were:

- CQ1: Are the believed circumstances true?
- CQ2: Assuming the circumstances, does the action have the stated consequences?
- CQ3: Assuming the circumstances and that the action has the stated consequences, will the action bring about the desired goal?
- CQ4: Does the goal realise the value stated?
- CQ5: Are there alternative ways of realising the same consequences?
- CQ6: Are there alternative ways of realising the same goal?
- CQ7: Are there alternative ways of promoting the same value?
- CQ8: Does doing the action have a side effect which demotes the value?
- CQ9: Does doing the action have a side effect which demotes some other value?
- CQ10: Does doing the action promote some other value?
- CQ11: Does doing the action preclude some other action which would promote some other value?
- CQ12: Are the circumstances as described possible?
- CQ13: Is the action possible?
- CQ14: Are the consequences as described possible?
- CQ15: Can the desired goal be realised?
- CQ16: Is the value indeed a legitimate value?
- CQ17: Will the other agent's prevent the stated consequences from being realised?

In order to provide a formal basis for this argumentation scheme, the problem scenario is modelled as an Action-based Alternating Transition System (AATS) [14]. Essentially, an AATS consists of a set of states and transitions between them representing the possible joint actions of the agents in a given state. These transitions are labelled with the values (motives) promoted by moving from the source state to the target state. In the AATS, the notion of a goal is identified with the new state, as, in our case study, it is assumed that all the features of the new state will be part of an agent's goal. The argumentation scheme and all of its associated critical questions are given formal definitions in terms of an AATS in [2].

Arguments following the above scheme are now based on a particular path from the current state to a new state in the AATS. As an agent can only perform one action in the given circumstances, arguments for different actions attack each other. The critical questions also point to possible counterarguments: a negative answer to, for example, CQ13 attacks any argument for the particular action. Once a set of arguments and the attacks between them has been generated on the basis of a specific AATS, the status of the arguments can be evaluated. To do this we form the arguments into a Value Based Argumentation Framework (VAF), introduced in [3]. A VAF is an extension of the argumentation frameworks (AF) of Dung [9]. In an AF an argument is admissible with respect to a set of arguments S if all of its attackers are attacked by some argument in S , and no argument in S attacks an argument in S . In a VAF an argument succeeds in defeating an argument it attacks only if its value is ranked as high, or higher, than the value of the argument attacked. In VAFs

audiences are characterised by their ordering of the values. Arguments in a VAF are admissible with respect to an audience A and a set of arguments S if they are admissible with respect to S in the AF which results from removing all the attacks which do not succeed with respect to the ordering on values associated with audience A . A maximal admissible set of a VAF is known as a Preferred Extension (PE).

One of the key elements that the above approach brings is the explicit distinction between arguments that can be made to reason about physical causal relations and arguments that can be made to reason about motivations and their priorities. The formal framework as described here allows us to predict what possible actions an agent would take given its values and an ordering on these values. This kind of reasoning is often employed in the verification stage, where it is argued that, given the circumstances, the suspect would have acted in a particular way to fulfill his or her motives. Since crimes are “by definition deviant behaviour, what people do not normally do” [13], the key to the plausibility of this argument is explaining why the accused acted in the aberrant fashion. While a typical agent would not have chosen the action using a “normal” value order, there is some value order on which it would be chosen: that the agent under suspicion in fact had this unusual ranking of values itself needs to be explained in order to make the agent’s choice of this action plausible. Consequently, when considered from the intentional stance, there are three elements to a plausible story:

1. a motive for the action,
2. an explanation of why the agent had this motive, and
3. an explanation of why the agent’s value order was such as to make this motive of sufficient importance.

As we describe in the next section, there are numerous different states that the defendant Claus von Bülow could have been trying to reach, and could well be motivated to reach, each of which require consideration in coming to a decision about the case. The critical questions draw out distinctions that need to be considered concerning the motivational preferences of the defendant, and the alternatives available for satisfying those motives.

In [4], we extended the above non-abductive view on practical reasoning with a separate argument scheme for abductive practical reasoning to allow for search and investigation cases. This abductive scheme makes it possible to infer an explanation for a set of circumstances in terms of a motivated action. As with the original scheme, there are critical questions associated with the abductive scheme that allow the elements of the explanation to be critiqued. Additionally, new critical questions were formulated which provided ways of asking for alternative explanations for the circumstances. The abductive scheme can be combined with the normal scheme, which enables us to reason both predictively and explanatory about motivated actions. In [4] we provided an example that represented a scenario where a person was killed by toppling off a bridge and

the question to be resolved was whether he jumped or was pushed. With the abductive scheme, several possible explanations for his fall were inferred and the normal scheme was then used to try to verify these explanations.² Formal definitions of the abductive argument scheme and its associated critical questions in terms of an AATS are given in [5].

In this paper, our focus is not on generating possible explanations, but rather to examine the plausibility of the contention that the defendant Claus von Bülow acted as alleged in the prosecution’s story. Here, we will simply adopt Thagard’s formulation of the story. Because the case can thus be considered a verification case, the reasoning will be mainly based on AS1 and its critical questions.

4 Case Study

We will use as our central case study the von Bülow case, formalised by Thagard in [11]. In 1980, Martha van Bülow, known as Sunny, a wealthy heiress and well-known socialite, went into a coma from which she eventually died in late 2008. In 1982 her husband, Claus von Bülow, was convicted of her attempted murder. In 1985 he was granted a new trial and acquitted.³ Thagard presents the case as an exercise in causal reasoning, notionally undertaken by the jurors, stating that ‘legal reasoning in trials such as those of Claus von Bülow’s can be characterized as inference to the best overall causal story’. By *best*, Thagard means most coherent according to his basic principles of explanatory coherence, which are implemented in terms of a connectionist algorithm.

We shall use Thagard’s formalisation as the basis for our representation of the case. Thagard divides his propositions into evidence, prosecution hypotheses, and defence hypotheses. Propositions can relate to one another either by one proposition contradicting another, or by one proposition explaining another. His approach is to use these relations in a connectionist model to determine a degree of explanatory coherence, propositions explaining one another increasing coherence (*excitatory links*) and propositions contradicting one another decreasing coherence (*inhibitory links*).

4.1 The First Trial

Figure 1 shows Thagard’s model of the first trial, where normal lines denote excitatory links and dotted lines denote inhibitory links. The evidence in the case was taken from a number of witnesses, as well as clearly observable facts. Thagard’s evidence is as follows. The short names in brackets are Thagard’s and used in Figure 1.

- *Observable Fact*: Sunny went into a coma (S-coma).

²Technically, we used did not use the full normal scheme in [4]. Rather, a critical question for the abductive scheme that represented the normal scheme in condensed form was used.

³There are a number of descriptions of the case on-line, for example there is a lengthy but readable account at www.trutv.com. It was filmed as *Reversal of Fortune* in 1990 starring Jeremy Irons as Claus and Glenn Close as Sunny.

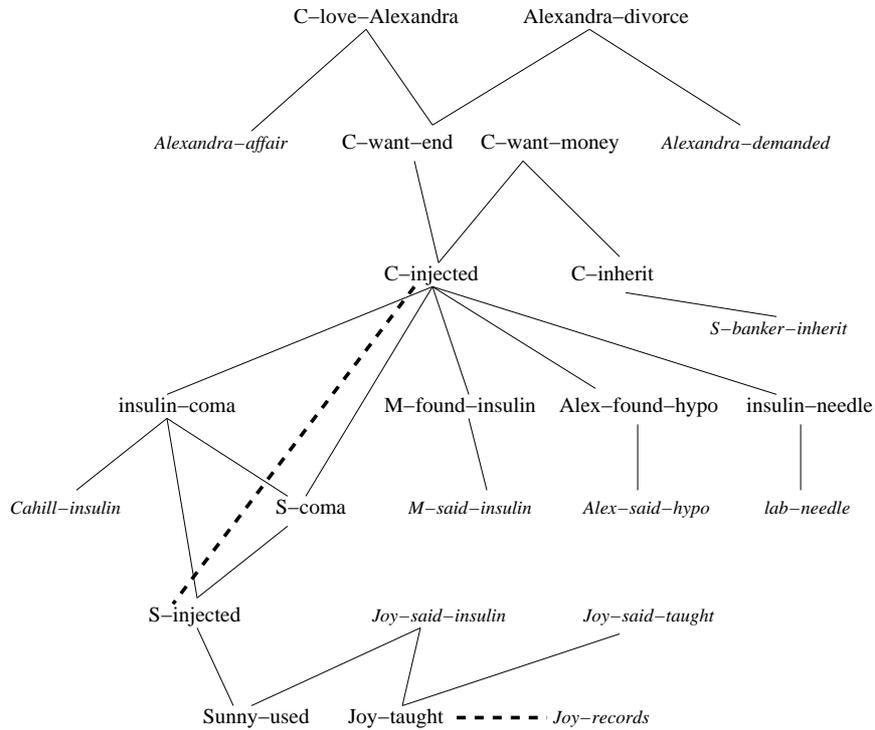


Figure 1: Model of first trial from Thagard.

- *Testimony from Maria* (Sunny's maid). Maria found insulin in Claus's bag (Maria-said-insulin)
- *Testimony from Alex* (Sunny's son by a previous marriage). Alex found a used hypodermic in Claus's bag (Alex-said-hypo).
- *Testimony from Cahill* (presumably Sunny's doctor). Insulin put Sunny in a coma (Cahill-insulin).
- *Testimony from Sunny's Banker*: Claus would inherit \$14 million if Sunny died (S-banker-inherit).
- *Testimony from Alexandra* (Claus's mistress). Alexandra was having an affair with Claus (Alexandra-affair) and she demanded that Claus get a divorce (Alexandra-demanded).
- *Testimony from Joy* (A personal trainer). Joy had taught Sunny many times (Joy-said-taught) and that Sunny had recommended insulin as a slimming aid (Joy-said-insulin).
- *Laboratory report*. There was insulin on the used hypodermic (lab-insulin)

- *Gym records.* Joy had hardly taught Sunny (Joy-records).

The prosecution argued that Joy’s evidence was not reliable, since it was contradicted by the gym records, but that the other testimonies should be believed. The evidence of Maria, Alex, Cahill and the lab report established that Claus had the means to inject Sunny. The testimonies of Alexandra and Sunny’s banker supplied a motivating story for Claus: that he loved Alexandra and wanted to end his marriage to Sunny, but that he also wanted to inherit the money (Claus had no independent means, and his opulent life style was supported by Sunny’s money). The defence story was based on Joy’s testimony and was that Sunny had injected herself with insulin as part of her eccentric health regime. These explanations are represented in Thagard as causal explanations and motive, but as can be seen above they are not treated differently in his model.

In our approach we will begin by focussing on the motive. We need first to identify some states and some transitions between them. We take the relevant features of the situation to be that Claus enjoyed a lavish lifestyle, but was dependent on Sunny for his money, that Claus was with Alexandra, and that Sunny was alive and conscious. We thus form our states from the five propositions:

- Dependent(Claus)
- LavishLifestyle(Claus)
- Together(Claus,Alexandra)
- Alive(Sunny)
- InsulinComa(Sunny)

Before Sunny went into a coma, the first four of these were true and the last was false, which we write as 11110.

Next we need to consider what actions might have been performed. Claus could leave Sunny, do nothing, or inject her with insulin. There are, of course, other possible actions, but these are all we need to consider here: alternative ways of attempting to kill Sunny are not of concern, and any other action can be taken as doing nothing since it has no relevance to the situation of interest. Sunny could do nothing or inject insulin. Alexandra could do nothing or leave Claus. We also need to consider that the effect of the injection could be one of three things: Sunny could control her weight, fall into a coma, or die. Finally we need to consider the police actions: they could charge Claus or accept accidental death. This gives us the transition diagram shown in Figure 2. Only the relevant states are shown.

We have labelled the transitions in the diagram with motives. Assuming that Claus loves Alexandra, or at least wishes that they stay together, his motivation of *love* (L) would be promoted where they remain together and demoted where they part. Claus is currently dependent on Sunny: he may well wish to be

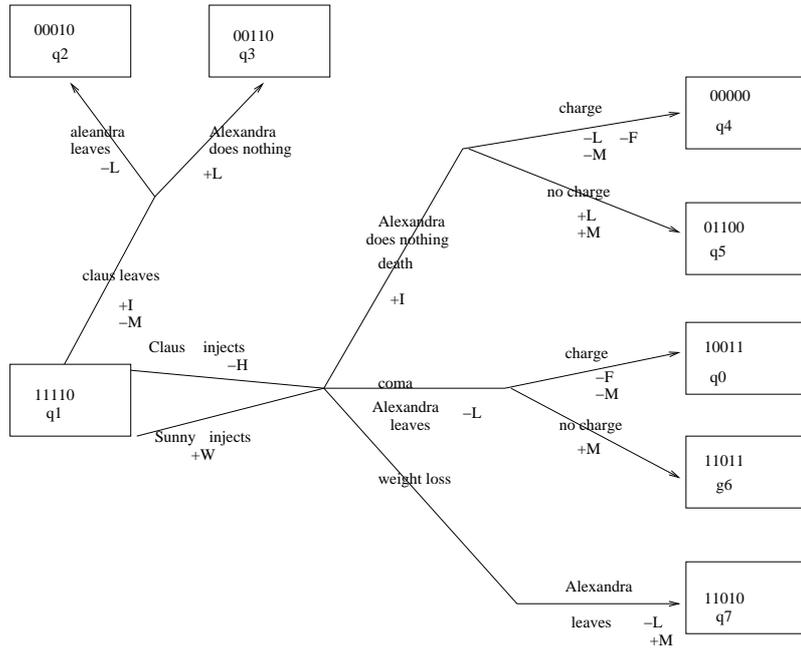


Figure 2: AATS for motive.

independent, so the motivation *independence* (I) is promoted when he leaves Sunny or she dies. But Claus was a notorious sybarite and so also might be motivated by a *lavish lifestyle* (M). Claus also is likely to value his *freedom* (F). We also have a motivation *honesty* (H) which is intended to embrace all the right reasons for obeying the law, such as respect for life, concern for reputation, moral reasons etc. Finally, Sunny was known to be concerned about her weight, and so controlling this (W) is a motivation for her.

The actual situation reached was q0. Obviously this situation was intended by neither Claus nor Sunny. The prosecution suggestion is that Claus was attempting to reach q5 by injecting Sunny with insulin, which would have satisfied three of his motivations, at the expense of Honesty, while the defence claims that Sunny was attempting to reach q7, which would have satisfied her motivation of weight loss.

The prosecution invites us to consider Claus in q1. He would have had three arguments based on AS1 for injecting Sunny, one for each of the values, and we will collapse these arguments into one:

CA1: In q1 I should inject to reach q5, which will promote L, I and M.

This is an instantiation of the standard practical reasoning scheme AS1 stated earlier. In the case the defence tried to tell a different story based on Sunny's position in q1, namely:

SA: In q1 I should inject to reach q7, which will promote W.

There are also numerous critical questions the defence can pose against CA1. Let us consider each of the critical questions in turn. Because we identify the goal (G) with the new circumstances (S), we will not consider CQ3, CQ6, or CQ15.

- CQ1: This would require the state to have been other than q1. But it is easy to establish Claus' financial dependence and lifestyle: Alexandra's testimony establishes their relationship, Sunny was indisputably alive and not comatose. Thus this question cannot sensibly be answered negatively.
- CQ2: As the AATS in figure 2 shows, Claus injecting Sunny could lead to any of q0, q4, q6 or q7 rather than q5. CQ2 applies to reaching q0 or q6 which suggests that the belief that the insulin would prove fatal is mistaken. The other possibilities rely on different choices by relevant agents, and so are covered under CQ17.
- CQ4: Assuming the facts of q1, the only dispute is whether killing Sunny would promote M, since it might be that her money went elsewhere. But Sunny's banker testifies that Claus would inherit a large sum.
- CQ5: Although there may have been other ways of killing Sunny, these were not proposed. It could have been argued, however, that had Claus intended to kill Sunny he could have chosen some surer and less detectable method to increase the chances of reaching q5. Since, however, this might have encouraged the belief that he did in fact want to reach q5, it would be dangerous to present such arguments, and we will not consider any based on this question.
- CQ7: Claus could have promoted L and I simply by leaving Sunny:
CA2 In q1 I should leave Sunny to reach q3 which will promote L and I.
- CQ8: None of these values are demoted in q5, so this question offers no assistance to Claus.
- CQ9: The action does demote H, which would provide Claus with a reason not to perform the action.
- CQ10: This question is important if we need to determine the particular motivation. In the current situation, however, Claus' reasons accrue rather than compete, and so we cannot use this question.
- CQ11: Again this question cannot be used, since we do not have any means of promoting other values in our problem formulation.
- CQ12: As the discussion of CQ1 implied, the circumstances were not only possible but in fact obtained.
- CQ13: Various bits of prosecution evidence were used to show that Claus had access to insulin, most notably the testimony of Alex and Maria.

- CQ14: Could injecting insulin have killed Sunny? The medical testimony would seem to establish this.
- CQ16: Are L, M and I really values that could motivate Claus? They seem common enough human desires to be acceptable.
- CQ17: It could be suggested that even if the insulin did kill Sunny, q5 would not be reached either because Alexandra would leave or because the police would charge Claus.

Based on these critical questions, the following counterarguments to CA1 can be constructed:

- ACQ1: Claus was not in love with Alexandra
- Aq0: Sunny would not die, and Claus would be charged, demoting F, M and L
- Aq6: Alexandra would leave anyway, demoting L
- Aq7: Sunny would neither die nor fall into a coma and Alexandra would leave, demoting L
- ACQ4: Claus would not inherit
- CA2: As stated above
- ACQ9: Injecting Sunny demotes H
- ACQ13: Claus had no access to insulin
- Aq6: Alexandra would leave anyway, demoting L
- Aq4: Claus would be charged with Sunny's murder, demoting F and M

These arguments in turn, however, have counterarguments of their own.

- TA: ACQ1 is attacked by Alexandra's testimony
- WC: Aq0 and Aq4 are attacked by the belief that Claus will not be charged.
- PF: Aq0, Aq6 and Aq7 are attacked by the belief that the injection will prove fatal
- TB: ACQ4 is attacked by the banker's testimony
- Aq2: CA2 is attacked by the belief that Alexandra will leave if Claus has no money (CQ2). This is itself attacked by AS, the belief that Alexandra will stay.
- MD: CA2 is attacked since it will demote M (CQ9)

- CA1: CA1 attacks CA2 through CQ7
- PT1: ACQ13 is attacked by Alex testimony
- PT2: ACQ13 is attacked by Maria’s testimony
- AS: Aq6 is attacked by the belief that Alexandra will not leave if Claus has money

We now arrange these arguments as a Value Based Argumentation Framework (VAF) [3]. The framework shown in Figure 3 is not a standard value-based framework: for clarity, and because here different motives reinforce one another rather than conflict, we have collected several arguments together. CA1 is thus associated with all three values promoted in q5. We have arguments from the various alternative states that might be reached, associated with the values they demote or fail to promote. We have arguments relating to the factual assumptions that Claus can make to rebut certain of the critical questions, that he won’t be charged and that Sunny will die. Finally we have the argument CQ13, which relates to means. When evaluating we give the argument the strength of the highest ranked value.

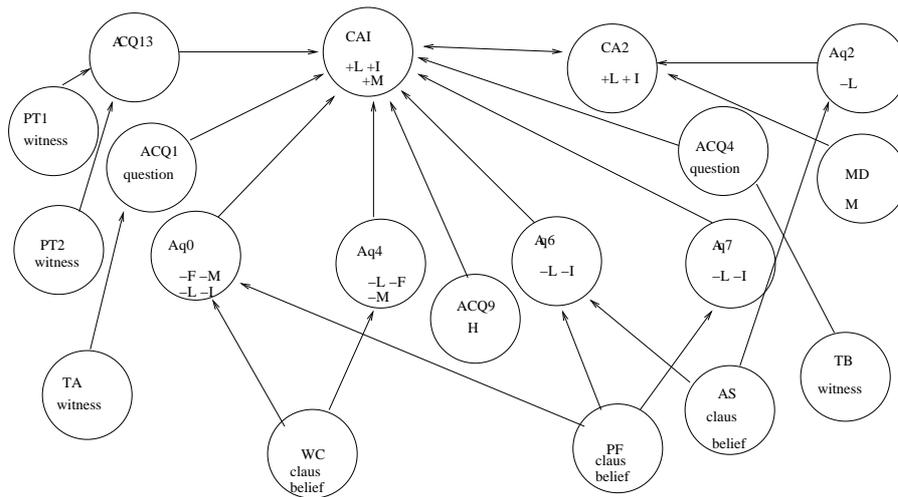


Figure 3: VAF for the scenario.

Now let us put ourselves in the position of the jury. The various testimonial evidence was unchallenged, and so TA, TB, PT1 and PT2 are all accepted. If this is so, ACQ1, ACQ13 and ACQ4 are all defeated. Now it is critical to judge that Claus would decide for CA1 rather than CA2. Suppose first that Claus believed that Alexandra would leave if he had no money. Then Claus would have a reason to reject CA2 and so accept CA1: only by killing Sunny could he keep Alexandra. But let us suppose that Claus believes that Alexandra will stay even if he has no money, so that Aq6 can be discounted. Now consider

some other possible value orders. We will consider L+I as a single value, since they are always together in the relevant arguments.

- $H > M > L+I$: here Claus cannot accept CAI because of respect for the law, but equally cannot accept CA2, as it would lose the money. Here Claus does not inject Sunny, but rather does nothing, accepting that Alexandra will leave.
- $H > L+I > M$: here Claus accepts CA2 and leaves Sunny
- $M > L+I$: here CA2 falls to MD, and so CAI is accepted, provided $M > H$.

From this a picture of the kind of person Claus must be to accept CA1 rather than CA2 emerges: he must be very motivated by a desire to preserve his lifestyle, but must also prize L over H. Unless this is so, he will either simply leave, or stay, depending on the importance of his feelings for Alexandra. It requires both M and L to be greater than H to accept CA1. But CA1 is still open to challenge in that he may not succeed in killing Sunny and escaping the consequences. Suppose that Claus did believe PF. Now only if he prized I over F would he discount the possibility of being charged, and it is hard to believe that he would have had this value order.

Thus the prosecution needs to convince the jury that Claus had certain preferences and certain beliefs. With regard to preferences, the prosecution needs to tell a story in which Claus ranks $M > L+I > H$. With respect to beliefs they would prefer for the jury to believe that Claus thought that Alexandra would leave if Claus had no money, since this defeats CA2 irrespective of Claus' preferences. It is, however, more plausible to think that Alexandra would stay with Claus if he left Sunny: she had asked him to leave Sunny and, as a successful TV star (*Dark Shadows*, a cult series which ran from 1966-71), did not really need money from Claus. This removes an objection against CA2, but Claus can still accept CA1 on the basis of the preferences just mentioned. The prosecution need not, therefore express any particular position on this. Next the jury must accept that Claus believed the injection would prove fatal. This is important, since otherwise CA1 cannot be acceptable, but the jury could very well accept this, since Claus is no medical expert, and such an injection might have proved fatal. The final question is whether Claus accepted WC, even though surrounded as he was by hostile servants and step children it seems inconceivable that he could have believed that he would not be the prime suspect. Since F is likely to be highly prized by Claus, the arguments that he would be deterred by the prospect of being caught are strong ones. Somehow, therefore the jury must also be convinced that he was sufficiently reckless or self confident to discount this possibility, and so be prepared to run the risk of being charged rather than not promote M⁴.

⁴At the trial no evidence was presented as to Alexandra's financial status. Had it been shown that she could have supported Claus in the style to which he was accustomed, M would become a lesser factor. However, now leaving Sunny would exchange dependence on

Suppose the jury did accept that Claus could have been motivated as suggested by the prosecution, and had the appropriate beliefs. The defence story was that Sunny had injected herself, and was supported by the testimony of Joy. This was challenged and fairly easily discredited. This left the prosecution story, showing means and motive as the only option, and it can be seen as plausible on a certain view of Claus's character. Evidently the jury took this view in the trial.

Thagard's approach is to use an artificial neural network that represents propositions by artificial neurons and represents coherence and incoherence relations by excitatory and inhibitory links between the units that represent the propositions. Acceptance or rejection of a proposition is represented by the degree of activation of the unit. As can be seen from Figure 1 the only inhibitor in the first Trial relates to the defence story, and so there is no real contest.

But let us consider, from our analysis of the arguments presented above, how the defence might have done better.

1. They might have been able to propose an alternative story more convincing than AS.
2. They might have been able to challenge the various pieces of prosecution testimony as to fact to provide some reason to reject TA, PT1 and PT2.
3. They might have disputed the motive. TB is particularly important in that without it M ceases to be a value promoted in CA1. Without M, it is no longer possible to attack CA2 on the grounds that it demotes M, or to defend CA1 by claiming that Claus preferred M to H. Thus without TB to support M as a motive CA2 looks like the more plausible choice for Claus, even if we continue to believe WC.
4. They might have argued that Claus could not have believed that he would not be suspected
5. They might have argued that Claus valued Honesty too much to commit the crime, or did not consider maintaining his lifestyle to be of importance

Some of these improvements were actually considered in the second trial, which we will consider now.

4.2 The Second Trial

In the second trial the defence was very different. This time Claus was represented by Alan Dershowitz, a Harvard Law School professor whose mastery of the appeals process is widely recognised, and who is among the best American appellate lawyers. First Dershowitz called expert witnesses to testify that Sunny's coma was not insulin induced, but was explained by a different story

one woman for dependence on another, as I would in this story have become the dominant motive, perhaps even worth risking his freedom for.

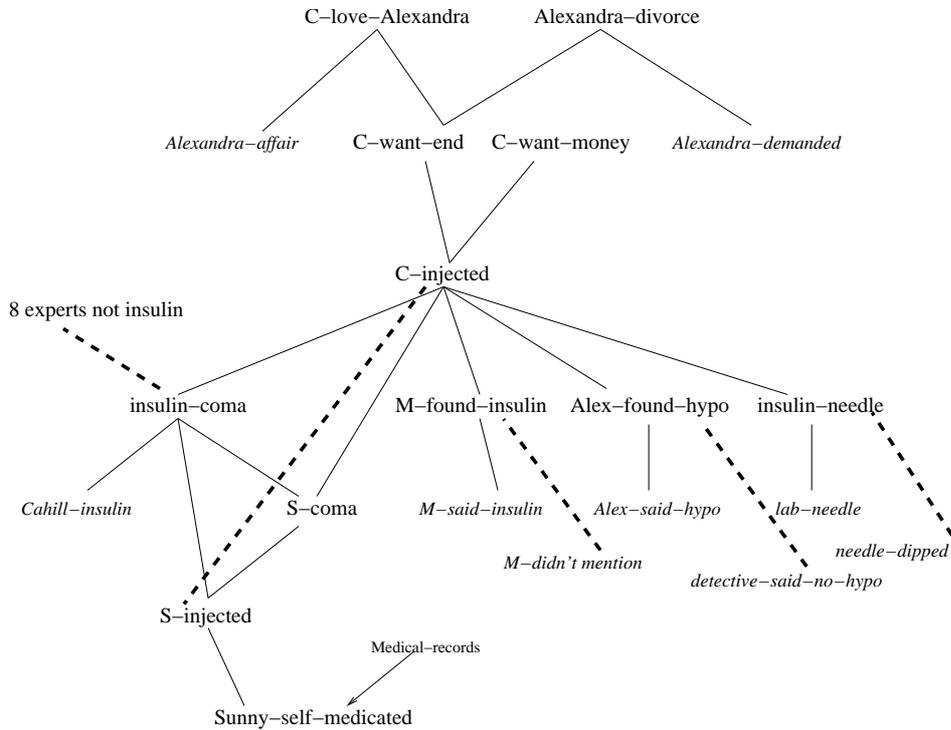


Figure 4: Model of second trial using Thagard's propositions.

relating to Sunny's many health problems and strange health behaviours (she had, for example, on occasion taken large amounts of aspirin and other drugs). This provides a much better basis for an alternative story than did the original story based on Joy's testimony. Instead of having to attribute to Sunny the unusual belief that insulin helped weight loss, and so requiring an implausible witness to establish this, they could rely on Sunny's known hypochondria and penchant for self-medication. This is improvement (1) from the last section.

Even though the defence was arguing that the coma was not insulin induced, they discredited the testimony of Maria (she had not initially mentioned the insulin in Claus's bag) and Alex (a detective who was with Alex when the bag was found had not seen any needles). Moreover they claimed that the hypodermic had been dipped in insulin rather than used to inject it. This casts doubt on PT1 and PT2, so reinstating ACQ13 (improvement 2 from the last section). Now even if the medical evidence that Sunny's coma was not insulin induced is rejected, doubt has been cast on whether Claus had the means to commit the crime.

Thirdly Dershowitz managed to get the banker's testimony about the money Claus would inherit excluded, so that M has to be excluded from Claus' motivation (improvement 3 above). Improvements 4 and 5 were not considered

necessary.

Now the prosecution case falls apart: Claus possibly did not have the means to inject insulin, and anyway insulin was not responsible for the coma. Moreover with the money element excluded from consideration, it would appear that Claus would have ended up independent but poor and this could have been achieved without risk by simply leaving Sunny, destroying the plausibility of Claus accepting CA1 rather than CA2. Reevaluating the VAF of Figure 3 without TA, PT1, PT2 and TB, leaves CA2 as the most acceptable argument. Similarly, Thagard's model of the second trial, shown in Figure 4, gives the defence the more coherent story, since inhibitory links now are associated with all the factual contentions of the defence, and the excitatory link from the banker's testimony has gone, and the inhibitory links casting doubt on the defence testimony have also disappeared. In contrast to the model of the first trial there are now so many inhibitors undermining the prosecution that the story cannot establish itself as coherent.

Although Thagard's approach, like ours, succeeds in giving the correct answer for the two cases, the differences between his models of the two trials is overwhelming. A fairer test would be to suppose that the banker's testimony was excluded from the first trial removing the second prong of the prosecution story about Claus's motivation. This would not make any difference to the explanatory coherence of the prosecution story using Thagard's model, since removing this excitatory link would not change the overall picture to any great extent. In our approach, however, where we can also consider matters from the *intentional* stance, it could make a considerable difference. If Claus was not going to inherit, if Sunny died and Claus was not charged, the state would be 00100 (q8) rather than q5 on the banker's testimony. This state q8 promotes only L and I, and this could have been achieved by leaving Sunny without demoting H, and without any risk to F. Thus, on our account the money element is essential rather than only supportive as in Thagard. The problem with Thagard's approach, from our perspective, is that although in his *representation* he distinguishes between causal explanations and motives, their *effect* on execution is identical, both acting as excitatory links in the same way. Thus, for Thagard, the excitatory link for Claus wanting to end his marriage causes Claus to inject Sunny is no different from that for the injection causing the coma. In particular there is no consideration of the alternative choices that Claus could have made, and no attempt to rationalise why Claus should have chosen to act on any particular motive, which is the case in our account.

5 Concluding Remarks

In this paper we have considered the role played by motive in persuading people of a person's guilt or innocence. Like Thagard we consider that the reasoning involved in such cases can be seen as inference to the best causal story. Unlike Thagard, however, we believe that it is important to consider the story not only from a physical stance, in which motives cause people to act in much the

same was as eating causes them to be overweight, but also from an intentional stance, so that they can be seen as rational agents making rational choices. Many aspects of the story depend on rational individuals making choices as to what to do, and while the existence of a motive does provide a reason for them to choose the action it is typically the case that there will be other motives, or alternative ways of acting on the motives. In order to believe that the story is the best explanation of the crime, we have to accept that the suspect could have believed that the action was the best choice in his particular situation. In order to provide this intentional stance we make use of a the approach to practical reasoning developed in [2] which enables us to identify the arguments that would have been available to the suspect, and to characterise the type of person who would have chosen as the suspect is alleged to have done.

This last aspect brings us on to the notion of character evidence, as discussed in [12]. We do not, however, discuss character evidence, other than to note that the role of character evidence is to establish that the suspect could plausibly have the ordering on values and motives required to make the choice he is alleged to have made. Nor do we discuss the admissibility of motive and character evidence, the central issue of [10] and touched on in [13]: rather we simply accept, like Thagard, that beliefs as to motive and character play an important role in the jury's assessment of the plausibility of the account of events presented, and do not consider how these beliefs are produced through the presentation of evidence.

Taking the case of Claus von Bülow as a case study enables direct comparison with Thagard's use of the case to demonstrate his own approach [11]. We believe that we have shown that like Thagard we are able to account for the decisions in the two trials. We would, however, claim the further advantage that our explanation is more transparent, being couched as explicit arguments, rather than resulting from the implicit operations of a connectionist algorithm. Moreover, although the evidence in the two trials was sufficiently different to make the outcomes relatively clear, in a case in which the evidence was rather less obviously stacked in favour of a particular side, the crucial questions can be readily identified from our explicit arguments. An example of this is the importance of the financial motive in the von Bülow case: while Thagard's model does not require this additional motive to produce a guilty verdict for the first trial, our account shows that without it Claus' alleged behaviour becomes far less plausible as he had easier and safer ways of achieving his other ends. This kind of scenario emphasises the need to consider the intentional stance. Whereas physical causes can be reasonably considered to have determinate effects, motives may or may not be acted on. Moreover, while causes *typically* give rise to their effects, so that it is when they do not that requires explanation in terms of some *abnormality* in the situation, the motives for a crime are typically not acted on: most people obey the law. Therefore some unusual feature of the situation which led this *particular* suspect to act on the motive in this *particular* case needs to be incorporated in a story if it is to be compelling.

References

- [1] K. Atkinson, T. Bench-Capon, and P. McBurney. Computational representation of practical argument. *Synthese*, 152(2):157–206, 2006.
- [2] K. Atkinson and T. J. M. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171(10–15):855–874, 2007.
- [3] T. Bench-Capon. Persuasion in practical argument using value based argumentation frameworks. *Journal of Logic and Computation*, 13(3):429–48, 2003.
- [4] F. J. Bex, T. J. M. Bench-Capon, and K. Atkinson. Did he jump or was he pushed? abductive practical reasoning. In *Legal Knowledge and Information Systems. JURIX 2008: The Twenty First Annual Conference*, pages 138–149. IOS Press, 2008.
- [5] F. J. Bex, T. J. M. Bench-Capon, and K. Atkinson. Did he jump or was he pushed? abductive practical reasoning. *Accepted for publication in Artificial Intelligence and Law*, 2009.
- [6] F. J. Bex, H. Prakken, and B. Verheij. Formalising argumentative story-based analysis of evidence. In *Proceedings of the Eleventh International Conference on AI and Law (ICAIL 2007)*, pages 1–10, 2007.
- [7] C. J. de Poot, R.J. Bokhorst, P.J. Van Koppen, and E.R. Muller. *Rechercheportret - Over dilemmas in de opsporing (Investigation portrait about dilemmas in investigation)*. Kluwer, Alphen a.d. Rijn, 2004.
- [8] D. Dennett. Intentional systems. In *Brainstorms*, pages 3–22. Harvester Press, 1979.
- [9] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
- [10] David P. Leonard. Character and motive in evidence law. *Loyola of Los Angeles Law Review*, 34:439–536, 2001.
- [11] P. Thagard. Causal inference in legal decision making: Explanatory coherence vs. bayesian networks. *Applied Artificial Intelligence*, 18:231–241, 2004.
- [12] D. Walton. *Character Evidence: An Abductive Theory*. Springer, Berlin, 2007.
- [13] D. Walton and B. Schafer. Arthur, George and the mystery of the missing motive: Towards a theory of evidentiary reasoning about motives. *International Commentary on Evidence*, 4(2):1–47, 2006.

- [14] M. Wooldridge and W. van der Hoek. On obligations and normative ability: Towards a logical analysis of the social contract. *Journal of Applied Logic*, 3:396–420, 2005.