On precise and correct qualitative probabilistic inference

Simon Parsons
Department of Computer and Information Science,
Brooklyn College,
City University of New York,
2900 Bedford Avenue,
Brooklyn, NY 11210, USA.

parsons@sci.brooklyn.cuny.edu

Abstract

In a recent paper I proposed a system for qualitative probabilistic reasoning, based on argumentation, and proved its correctness with respect to probability theory. This system was flawed. In particular, it failed to take proper account of d-separation, and so can give erroneous results in certain cases. This paper identifies some of the problems caused by this flaw, examines their extent, and then fixes the flaw. While the main thrust of the paper is to overcome this flaw, the discussion of the problems caused by the flaw exposes some general issues in qualitative probabilistic reasoning.

1 Introduction

In the last few years there have been a number of attempts to build systems for reasoning under uncertainty that are of a qualitative nature—that is they use qualitative rather than numerical values, dealing with concepts such as increases in belief and the relative magnitude of values. Between them, these systems address the problem of reasoning in situations in which knowledge is uncertain, but in which there is a limited amount of numerical information quantifying the degree of uncertainty. One class of these systems are systems of abstraction. In systems of abstraction, the focus is mainly on modelling how the probability of a hypothesis changes when evidence is obtained and there is no need to commit to exact probability values. They thus provide an abstract version of probability theory, known as qualitative probabilistic networks (QPNs), which ignores the actual values of individual probabilities but which is nevertheless sufficient for planning [18], explanation [2] and prediction [11] tasks. Another class are systems of argumentation. Systems of argumentation are based on the idea of constructing logical arguments for and against formulae, establishing the overall validity of such formulae by assessing the persuasiveness of the individual arguments. Systems of argumentation have been applied to problems such as diagnosis, protocol management and risk assessment [5], as well as handling inconsistent information [1], and providing a framework for default reasoning [4, 8, 14].

In a recent paper [10] I described a hybridisation of the argumentation and abstraction approaches by introducing a logical system for reasoning about how probabilities change, called the qualitative probabilistic reasoner ($\mathcal{QPR}$). The input to this system is a set of logical formulae describing probabilistic relationships between variables, and information about how the probabilities of particular formulae change. In [10] I showed that the system can establish exactly those changes in probability in other formulae that are sanctioned by probability theory. However, there is a flaw in $\mathcal{QPR}$ as originally defined. The main contribution of this paper is to identify the flaw, discuss its consequences, fix it. Since the effects of the flaw hinge on the interaction between different kinds of qualitative probabilistic information, this paper also makes a more general contribution to the study of qualitative probabilistic inference in identifying this interaction.

The structure of the paper is as follows. The next two sections, Sections 2 and 3 introduce a cut down version of $\mathcal{QPR}$, termed $\mathcal{QPR}_\mathcal{C}$, which is sufficient to illustrate the flaw. Section 4 then identifies the flaw $\mathcal{QPR}/\mathcal{QPR}_\mathcal{C}$ and elaborates on its consequences, before Section 5 shows how it may be solved.

2 The logical language

The system introduced here is basically $\mathcal{QPR}$ from [10] without synergies, and with no proof rules for evidential or intercausal reasoning. Thus $\mathcal{QPR}_\mathcal{C}$ is a version of $\mathcal{QPR}$ which is only capable of reasoning in a causal direction (hence the name).

2.1 Basic concepts

We start with a set of atomic propositions $\mathcal{L}$. We also have a set of connectives $\{\neg, \wedge, \rightarrow\}$, and the following set of rules for building the well-formed formulae (*wff* s) of the language.

1. If $l \in \mathcal{L}$ then l is a simple well-formed formula (*swff*).
2. If l is an *swff*, then $\neg l$ is an *swff*.
3. If l and m are *swff* s, then $l \wedge m$ is an *swff*.
4. If l and m are *swff* s then $l \rightarrow m$ is an implicational well-formed formula (*iwff*).
5. The set of all *wff* s is the union of the set of *swff* s and the set of *iwff* s.

There is an important point that should be noted about the connectives which go to make up these formulae—that $\rightarrow$ does not represent material implication. Instead it represents a constraint on the conditional probabilities relating the formulae it connects. Such constraints have exactly the form of the constraints embodied in the qualitative influences of QPNs, albeit for variables with binary values, and their precise semantics is given below.

The set of all *wff*s that may be defined using $\mathcal{L}$, may then be used to build up a database Δ where every item $d \in \Delta$ is a triple $(i : l : s)$ in which i is a token uniquely identifying the database item (for convenience we will use the letter 'i' as an anonymous identifier), l is a *wff*, and s gives information about the probability of l. In particular we take triples $(i : l : \uparrow)$ to denote the fact that $\Pr(l)$ increases, and similar triples $(i : l : \downarrow)$, to denote the fact that $\Pr(l)$ decreases. Triples $(i : l : \leftrightarrow)$, denote the fact that $\Pr(l)$ is known to neither increase nor decrease. It should be noted that the triple $(i : l : \uparrow)$ indicates that $\Pr(l)$ either goes up, or does not change—this inclusive interpretation of the notion of "increase" is taken from QPNs—and of course a similar proviso applies to $(i : l : \downarrow)$. Since we want to reason about changes in belief which equate to the usual logical notion of proof, we also consider increases in belief to 1 and decreases in belief to 0, indicating these by the use of the symbols $\Uparrow$ and $\Downarrow$, and the values 1 and 0. The meaning of a triple $(i : l : \Uparrow)$ is that the probability of l becomes 1 if it is not 1 already, $(i : l : \Downarrow)$ means that the probability of l becomes 0 if it is not already. $(i : l : 1)$ means that the probability of l is 1 and $(i : l : 0)$ means that the probability of l is 0. We also have triples $(i : l : \updownarrow)$ which indicate that the change in $\Pr(l)$ is unknown. In addition, for reasons which will become clear later, we need a symbol to denote a probability whose value is not known (as distinct from a change in probability whose value is not known). This symbol will be $\wr$, so the triple $(i : l : \wr)$ means that the value of $\Pr(l)$ is unknown, but is known not to change. While this profusion of symbols might seem baroque, it is unfortunately necessary in order to distinguish the different aspects of qualitative probabilistic reasoning.[1]

2.2 Non-material implication

As mentioned above, $\rightarrow$ does not represent material implication but a connection between the probabilities of antecedent and consequent. This is the key to understanding the system. We take *iwff*s, which we will also call "implications",

[1] Dealing with categorical influences is the root cause of this profusion of symbols (compare the set of signs used here with those in [13] for example). Broadly speaking, the signs used in QPNs ($\uparrow$, $\downarrow$, $\updownarrow$ and $\leftrightarrow$ in the notation used here) represent first derivatives of probability values with respect to evidence (a point expanded on at length in [11]). While the only implications we have are non-categorical, we are only dealing with derivatives and these are the only values we need to consider. Once we introduce catagorical influences, we also introduce 1 and 0, landmark values in the terminology of qualitative reasoning [7], which are probabilities that have not been differentiated. The remaining signs arise from a need to have a set of values that is closed under the operations carried out on them during inference. Of course, there is some interconnection between the two sets of values (which is not there in regular qualitative reasoning) since knowing that a probability is 1, 0 or $\wr$ tells us that there it has a zero derivative, and so it has value $\leftrightarrow$.

to denote that the antecedent of the *iwff* has a probabilistic influence on the consequent. Thus we are not concerned with the probability *of* the *iwff*, but what the *wff* says *about* the probabilities of its antecedent and consequent. More precisely we take the triple $(i : a \rightarrow c : +)$ to denote the fact that:

$$\Pr(c|a, X) \geq \Pr(c|\neg a, X)$$

for all X for which there is a triple $(i : X \rightarrow c : s)$ (where s is any sign). The effect of the X in this inequality is to ensure that the restriction holds whatever is known about formulae other than c and a—whatever the probabilities of a and c, the constraint on the conditional probabilities holds. Similarly the triple $(i : a \rightarrow c : -)$ denotes the fact that:

$$\Pr(c|a, X) \leq \Pr(c|\neg a, X)$$

again for all X for which there is a triple $(i : X \rightarrow c : s)$. It is possible to think of an implication $(i : a \rightarrow c : +)$ as meaning that there is a constraint on the probability distribution over the formulae c and a such that an increase in the probability of a entails an increase in the probability of c, and an implication $(i : a \rightarrow c : -)$ means that there is a constraint on the probability distribution over the formulae c and a such that an increase in the probability of a entails a decrease in the probability of c. We do not make much use of triples such as $(i : c \rightarrow a : 0)$[2] since they have no useful effect but include them for completeness—$(i : c \rightarrow a : 0)$ indicates that:

$$\Pr(c|a, X) = \Pr(c|\neg a, X)$$

for all X for which there is a triple $(i : X \rightarrow c : s)$, and so denotes the fact that $\Pr(c)$ does not change when $\Pr(a)$ changes. We also have implications such as $(i : a \rightarrow c : ?)$ which denotes the fact that the relationship between $\Pr(c|a, X)$ and $\Pr(c|\neg a, X)$ is not known, so that if the probability of a increases it is not possible to say how the probability of c will change.

With this interpretation, implications correspond to qualitative influences in QPNs. Just as in QPNs, we often take implications to be causally directed, by which we mean that the antecedent is a cause of the consequent, and enforce the condition that chains of these directed links do not form a cycle. Thus we can consider every set of implications to have an associated QPN, where each arc in the QPN maps to an implication.

This simple picture is complicated because we have *categorical* implications which allow formulae to be proved true or false. In particular, an implication $(i : a \rightarrow c : {++})$ indicates that when a is known to be true, then so is c. Thus it denotes a constraint on the probability distribution across a and c such that if $\Pr(a)$ becomes 1, then so does $\Pr(c)$. This requires that:

$$\Pr(c|a, X) = 1$$

[2]As a result we will not worry about the possibility of confusing $(i : l \rightarrow m : 0)$ with $(i : l : 0)$ where l and m are *swff*s.

for all X for which there is a triple $(i : X \rightarrow c : s)$ [9]. Note that this type of implication also conforms to the conditions for implications labelled with + (and so may be considered as a more precise specialisation of an implication labelled with a +), and that if $\Pr(c|\neg a, X) = 1$ as well, then $\Pr(c)$ is always equal to $\Pr(a)$. Similarly, a probabilistic interpretation of an implication $(i : a \rightarrow c : --)$ which denotes the fact that if a is true then c is false, requires that:

$$\Pr(c|a, X) = 0$$

for all X for which there is a triple $(i : X \rightarrow c : s)$. The conditions imposed on the conditional values by these implications suggest the existence of a further pair of types of categorical implication which are symmetric to those already introduced. We have an implication $(i : a \rightarrow c : -+)$ which denotes the constraint:

$$\Pr(c|\neg a, X) = 1$$

for all X for which there is a triple $(i : X \rightarrow c : s)$, and an implication $(i : a \rightarrow c : +-)$ which denotes the constraint:

$$\Pr(c|\neg a, X) = 0$$

for all X for which there is a triple $(i : X \rightarrow c : s)$.

As mentioned above, the full system $\mathcal{QPR}$ allows for the representation of probabilistic synergies, in particular product synergies [2, 19]. It is also possible to include additive synergies [2, 18] and utilities [13] into this kind of system.

3 The proof theory

For the language introduced in the previous section to be useful we need to give a mechanism for taking sentences in that language and using them to derive new sentences. In particular we need to be able to take sentences describing changes in probability in particular formulae and use these to establish changes in probability in other formulae. This is done using the consequence relation $\vdash_{QP}$, part of which is defined in Figure 1. The definition is in terms of Gentzen-style proof rules where the antecedents are written above the line and the consequent is written below. The consequence relation operates on a database of the kind of triples introduced in the previous section and derives *arguments* about formulae from them. The concept of an argument is formally defined as follows:

Definition 1 *An* argument *for a well-formed formula p from a database Δ is a triple (p, G, s) such that $\Delta \vdash_{QP} (p, G, s)$*

The *sign s* of the argument denotes something about the probability of p while the *grounds G* identify the elements of the database used in the derivation of p.

To see how the idea of an argument fits in with the proof rules in Figure 1, consider the following example.

C-rules

$$\text{Ax}\frac{}{\Delta \vdash_{QP} (St, \{i\}, Sg)}\ (i : St : Sg) \in \Delta$$

$$\wedge\text{-E1}\frac{\Delta \vdash_{QP} (St \wedge St', G, Sg)}{\Delta \vdash_{QP} (St, G, \mathsf{conj}_{\mathsf{elim}}(Sg))}$$

$$\wedge\text{-E2}\frac{\Delta \vdash_{QP} (St \wedge St', G, Sg)}{\Delta \vdash_{QP} (St', G, \mathsf{conj}_{\mathsf{elim}}(Sg))}$$

$$\wedge\text{-I}\frac{\Delta \vdash_{QP} (St, G, Sg) \quad \Delta \vdash_{QP} (St', G', Sg')}{\Delta \vdash_{QP} (St \wedge St', G \cup G', \mathsf{conj}_{\mathsf{intro}}(Sg, Sg'))}$$

$$\neg\text{-E}\frac{\Delta \vdash_{QP} (\neg St, G, Sg)}{\Delta \vdash_{QP} (St, G, \mathsf{neg}(Sg))}$$

$$\neg\text{-I}\frac{\Delta \vdash_{QP} (St, G, Sg)}{\Delta \vdash_{QP} (\neg St, G, \mathsf{neg}(Sg))}$$

$$\rightarrow\text{-E}\frac{\Delta \vdash_{QP} (St, G, Sg) \quad \Delta \vdash_{QP} (St \rightarrow St', G', Sg')}{\Delta \vdash_{QP} (St', G \cup G', \mathsf{imp}_{\mathsf{elim}}(Sg, Sg'))}$$

Figure 1: The causal part of the consequence relation $\vdash_{QP}$

Example 1. We have a database which denotes the fact that the proposition "*premise*" has a probability which increases to 1, and that there is a relation between the proposition *premise* and the proposition "*conclusion*" such that if the probability of *premise* becomes 1, so does the probability of *conclusion*. This database is denoted:

$$\begin{array}{ll} (f1 : \mathit{premise} : \Uparrow) & \Delta_1 \\ (r1 : \mathit{premise} \rightarrow \mathit{conclusion} : ++) & \end{array}$$

From the database, by application of Ax it is possible to establish two simple arguments:

$$\Delta_1 \vdash_{QP} (\mathit{premise}, \{f1\}, \Uparrow)$$

denoting that on the basis of $f1$ we can infer that the probability of *premise* either increases to, or remains at, 1, and

$$\Delta_1 \vdash_{QP} (\mathit{premise} \rightarrow \mathit{conclusion}, \{r1\}, ++)$$

denoting that on the basis of $r1$ we can infer that there is a connection between *premise* and *conclusion* such that if the probability of the former increases to (or is) 1, then the probability of the latter increases to (or is) 1. Now, taking these two and applying $\rightarrow$-E, it is possible to build the argument:

$$\Delta_1 \vdash_{QP} (\mathit{conclusion}, \{r1, f1\}, \Uparrow)$$

	$++$	$+-$	$+$	0	$-$	$-+$	$--$	$?$
1	1	$\wr$	$\wr$	$\wr$	$\wr$	$\wr$	0	$\wr$
$\Uparrow$	$\Uparrow$	$\uparrow$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\downarrow$	$\Downarrow$	$\updownarrow$
$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\updownarrow$
$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$	$\leftrightarrow$
$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\leftrightarrow$	$\uparrow$	$\updownarrow$	$\uparrow$	$\updownarrow$
$\Downarrow$	$\downarrow$	$\Downarrow$	$\downarrow$	$\leftrightarrow$	$\uparrow$	$\Uparrow$	$\uparrow$	$\updownarrow$
0	$\wr$	0	$\wr$	$\wr$	$\wr$	1	$\wr$	$\wr$
$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\leftrightarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$
$\wr$	$\wr$	$\wr$	$\wr$	$\wr$	$\wr$	$\wr$	$\wr$	$\wr$

Table 1: Implication elimination $\mathsf{imp_{elim}}$

since applying $\mathsf{imp_{elim}}$ to $\Uparrow$ and $++$ yields $\Uparrow$ (as we will see in a little while). Thus from the database it is possible to build an argument for the probability of *conclusion* becoming (or being) 1. □

In order to apply the proof rules to build arguments, it is necessary to supply the functions used in Figure 1 to combine signs. A full definition of $\mathcal{QPR_C}$ thus requires the functions $\mathsf{conj_{elim}}$, $\mathsf{conj_{intro}}$, neg, and $\mathsf{imp_{elim}}$. However, since all are given in [10], and only $\mathsf{imp_{elim}}$ is used in this paper, this is the only function which will be given here.

The function $\mathsf{imp_{elim}}$ is used to establish the sign of formulae generated by the rule of inference $\rightarrow$-E. This means that $\mathsf{imp_{elim}}$ is used to combine the change in probability of a formula a, say, with the constraint that the probability of a imposes upon the probability of another formula c. Since this constraint is expressed in exactly the same way as qualitative influences are in QPNs, $\mathsf{imp_{elim}}$ performs the same function as $\otimes$ [18], and is merely an extension of it.

Definition 2 *The function* $\mathsf{imp_{elim}} : Sg \in \{1, \Uparrow, \uparrow, \leftrightarrow, \downarrow, \Downarrow, 0, \updownarrow, \wr\} \times Sg' \in \{++, +-, +, 0, -, -+, --, ?\} \mapsto Sg'' \in \{1, \Uparrow, \uparrow, \leftrightarrow, \downarrow, \Downarrow, 0, \updownarrow, \wr\}$ *is specified by Table 1.*

It is worth noting that Table 1 not only deals with the combination of changes in probability, such as $\uparrow$, with probabilistic constraints, but also gives the results of combining constraints and actual values like 1 and 0. Most of the time these combinations give a value of $\wr$, which denotes a probability that does not change and whose unchanging value is unknown.

4 The unbearable incorrectness of inference

As introduced in [10], $\mathcal{QPR}$ had three distinct sets of proof rules. One set, reproduced here, permitted reasoning in a causal direction, that is in the direction of the implications. Another set permitted reasoning in an evidential direction, that is in the opposite direction to the implications, and the third set permitted reasoning with synergies capturing intercausal reasoning [19]. The soundness and completeness proofs for $\mathcal{QPR}$ were given by considering first the causal rules, then the causal and evidential rules, and finally all three sets of rules together. Here we examine the causal rules, because it is here that the flaw we are primarily interested in resides.

4.1 The flaw and its consequences

Now, once we have applied the proof rules we find we have several arguments for a given proposition p. Thus we have an *argument set* for the proposition:

Definition 3 *The* argument set $\mathbf{A}_p^\Delta$ *for a proposition p from a database Δ is the set of all arguments for p which may be constructed from Δ:*

$$\mathbf{A}_p^\Delta = \{(p, G_i, Sg_i) \mid \Delta \vdash_{QP} (p, G_i, Sg_i)\}$$

Each of these arguments has a sign that summarises how the probability of p changes according to the information in that argument. Typically we are interested in the effect of the information in all the arguments. To establish this, we introduce a *flattening* function flat which combines arguments by mapping from a set of arguments A_p^Δ to the supported formula p and some overall change in probability:

$$\mathsf{flat} : \mathbf{A}_p^\Delta \mapsto \langle p, v\rangle$$

where v is the result of a suitable combination of the signs of the arguments. Now, because the effect of each implication is defined to occur whatever other arguments are formed (this is a result of the constraint imposed on the conditional probabilities by the implications), all combinations are completely local, and the structure of the arguments may be disregarded when flattening. As a result, v is simply calculated as:

$$v = \bigoplus_i s_i$$

for all $(p, G_i, s_i) \in \mathbf{A}_p^\Delta$ where $\oplus$ is an extended version of the qualitative addition function used by QPNs, defined as follows:

Definition 4 *The function* $\oplus : Sg \in \{1, \Uparrow, \uparrow, \leftrightarrow, \downarrow, \Downarrow, 0, \updownarrow, \wr\} \times Sg' \in \{1, \Uparrow, \uparrow, \leftrightarrow, \downarrow, \Downarrow, \updownarrow, \wr\} \mapsto Sg'' \in \{\Uparrow, \uparrow, \leftrightarrow, \downarrow, \Downarrow, \updownarrow, \wr\}$ *is specified by Table 2. Blank spaces represent impossible combinations.*

The blank spaces in Table 2 are an important feature which deserve some explanation. They arise as the result of the combination of a $\Uparrow$ and a $\Downarrow$—an increase to a probability of 1 and a decrease to a probability of 0—or the combination of $\Uparrow$ or $\Downarrow$ with 0 or 1 respectively. These are simply incompatible in the sense that it isn't possible to define a probability distribution which will allow this behaviour (something that is proved in Theorem 11 below).

With flat and $\oplus$ established we can give the overall procedure for determining the change in probability of a formula p in which we are interested. This is:

1. Add a triple $(i : q : s)$ for every formula q whose change in probability is known.

	1	$\Uparrow$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\Downarrow$	0	$\updownarrow$	$\wr$
1	1	1	1	1	1			1	1
$\Uparrow$	1	$\Uparrow$	$\Uparrow$	$\Uparrow$	$\Uparrow$			$\Uparrow$	$\Uparrow$
$\uparrow$	1	$\Uparrow$	$\uparrow$	$\uparrow$	$\updownarrow$	$\Downarrow$	0	$\updownarrow$	$\uparrow$
$\leftrightarrow$	1	$\Uparrow$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\Downarrow$	0	$\updownarrow$	$\leftrightarrow$
$\downarrow$	1	$\Uparrow$	$\updownarrow$	$\downarrow$	$\downarrow$	$\Downarrow$	0	$\updownarrow$	$\downarrow$
$\Downarrow$			$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$	0	$\Downarrow$	$\Downarrow$
0			0	0	0	0	0	0	0
$\updownarrow$	1	$\Uparrow$	$\updownarrow$	$\updownarrow$	$\updownarrow$	$\Downarrow$	0	$\updownarrow$	$\updownarrow$
$\wr$	1	$\Uparrow$	$\uparrow$	$\leftrightarrow$	$\downarrow$	$\Downarrow$	0	$\updownarrow$	$\wr$

Table 2: Flattening flat

2. Build $\mathbf{A}_p^{\Delta}$ using the C-rules.
3. Flatten this set to $\mathsf{Flat_c}(\mathbf{A}_p^{\Delta})$ where $\mathsf{Flat_c}(\mathbf{A}_p^{\Delta}) = \mathsf{flat}(\mathbf{A}_p^{\Delta})$.

Flattening is described in this way to allow for different flattening mechanisms to be used for different kinds of reasoning while still using flat (see [10] for details).

We can now see how problems arise and establish what the flaw in $\mathcal{QPR}/\mathcal{QPR_C}$ is. This can be done through some examples. Consider the following:

Example 2. The following clauses represent the fact that a has a positive influence on b, and b has a positive influence on c.

$$\begin{array}{ll} (r1 : a \rightarrow b : +) & \Delta_2 \\ (r2 : b \rightarrow c : +) & \end{array}$$

Now, consider we have evidence that a is suddenly observed to be true, so that the triple $(f1 : a : \Uparrow)$ is added to the database, it is possible to build the following argument concerning $\Pr(c)$ using $\mathcal{QPR_C}$:

$$\Delta_2 \quad \vdash_{QP} \quad (c, \{f1, r1, r2\}, \ \uparrow)$$

This is built by combining $f1$ and $r1$ using $\rightarrow$-E, and then using the result of this inference with $r2$ using $\rightarrow$-E again. This argument may then be flattened to give the pair $\langle c, \ \uparrow \rangle$. $\square$

This is entirely correct as one would hope from such a simple example. However, consider what happens in the following small variation on the example:

Example 3. Here b is known to be true:

$$\begin{array}{ll} (r1 : a \rightarrow b : +) & \Delta_3 \\ (r2 : b \rightarrow c : +) & \\ (f2 : b : 1) & \end{array}$$

Now, with the same additional information as before, we get two arguments about $\Pr(c)$ using $\mathcal{QPR_C}$:

$$\begin{array}{lll} \Delta_3 & \vdash_{QP} & (c, \{f1, r1, r2\},\ \uparrow) \\ \Delta_3 & \vdash_{QP} & (c, \{f2, r2\},\ \wr) \end{array}$$

and again the two arguments may then be flattened to give the pair $\langle c,\ \uparrow\rangle$. $\square$

This second example is not entirely correct in that it does not make as precise a prediction as is possible using probability theory.

The problem is this. $\mathcal{QPR_C}$ predicts that the probability of c will either increase or remain the same—this is the inclusive reading of $\uparrow$ that is standard in the QPN literature and which stems from the inequality in the denotation of the "$\rightarrow$" symbol. However, since b is known to be true, any subsequent probabilistic propagation from a to c is blocked. In the terminology of Bayesian networks, c is d-separated ([6], pages 7–14) from a. As a result, probability theory tells us that irrespective of changes in the probability of a, the probability of c will not change, and this is not respected in $\mathcal{QPR_C}$. $\mathcal{QPR_C}$ allows for the fact that the probability of c may not increase, since it hedges its bets somewhat with the prediction of $\uparrow$, and so makes a sound prediction. However, there is definitely a flaw here that results from $\mathcal{QPR_C}$ not handling d-separation in causal reasoning.

In fact, the situation is worse than this first analysis suggests. It is possible for $\mathcal{QPR_C}$ to actually give incorrect results. Consider this variation on the last example.

Example 4. Here b is known to be true:

$$\begin{array}{ll} (r1 : a \rightarrow b : ++) & \Delta_4 \\ (r2 : b \rightarrow c : ++) & \\ (r3 : a \rightarrow d : +) & \\ (r3 : d \rightarrow c : -) & \\ (f2 : b : 1) & \end{array}$$

Now, with the additional information that $(f1 : a : \Uparrow)$, we get three arguments about $\Pr(c)$ using $\mathcal{QPR_C}$:

$$\begin{array}{lll} \Delta_3 & \vdash_{QP} & (c, \{f1, r1, r2\},\ \Uparrow) \\ \Delta_3 & \vdash_{QP} & (c, \{f2, r2\},\ \wr) \\ \Delta_3 & \vdash_{QP} & (c, \{f1, r3, r4\},\ \downarrow) \end{array}$$

and again the two arguments may then be flattened to give the pair $\langle c,\ \Uparrow\rangle$, whereas the correct answer, were d-separation taken into account,would be $\langle c,\ \downarrow\rangle$. $\square$

This is the error alluded to in the title of this section.

Now, this particular error can only occur if the chain of inferences that should be blocked by d-separation is all categorical (otherwise the worst that could

1	Probability of 1
0	Probability of 0
$\wr$	Unknown probability, value known not to change
$\leftrightarrow$	No change in probability
$\Uparrow$	Probability increases to, or remains at, 1
$\Downarrow$	Probability increases to, or remains at, 0
$\uparrow$	Increase in probability, or no change
$\downarrow$	Decrease in probability, or no change
$\updownarrow$	Increase, decrease, or no change in probability

Table 3: The full set of predictions possible in $\mathcal{QPR}_{\mathcal{C}}$.

happen is the first argument has sign $\uparrow$ and flattening produces $\updownarrow$ rather than $\downarrow$, which is imprecise but not wrong), and if there is "old evidence" (evidence that is already in the system) in order to generate a need for d-separation to be taken into account in the first place. Since all systems that stem from QPNs are basically intended to only cope with single pieces of evidence (certainly that seems to be the case for Wellman's original formulation [18]) whenever we deal with more than one piece of evidence we are pushing the limits of what is possible [16]. As a result, this error is not surprising. However, it is worth fixing it.

4.2 The extent of the problems

Before attempting to fix the flaw, it is worth examining the extent of its consequences—identifying when it causes incorrect inferences to be drawn. In the lst example above, we can see the existence of such an incorrect inference, but can we obtain some results that formally circumscribe this kind of problem?

Following [11], it is possible to define the following concepts which allow us to get an idea of the extent of the problem.

Definition 5 *If applying* $\vdash_{QP}$ *to a database* Δ *generates an argument* (p, G, π), *then* π *is a* prediction *about the change in probability of* p.

Now, as we have seen in Section 2, there is a whole menagerie of different predictions, replicated in Table 3. The top set of predictions are essentially predictions that the probability of the given proposition will not change. The next set of values are predictions that either there will be a change in one direction, increase or decrease, or no change. The final value is a prediction that there will be an increase, decrease, or no change—it is less a prediction than an admission that it is impossible to tell how the value will change.

Now, the important point here is that the higher values in the table make more precise predictions about changes, in the sense that knowing the value will not change is more precise than knowing that it might increase (will either increase or not change). Thus there is an order over the possible predictions in terms of their precision:

Definition 6 *The set of predictions* $\Pi_0 = \{1, 0, \wr, \leftrightarrow\}$ *is the set of* no change *predictions, the set* $\Pi_1 = \{\Uparrow, \Downarrow, \uparrow, \downarrow\}$ *is the set of* change *predictions, and* $\Pi_2 = \{\updownarrow\}$ *is the set of* vacuous *predictions.*

Definition 7 *Given two predictions* π *and* π'*, from sets* Π_i *and* Π_j *respectively,* π *is* more precise *than* π' *if* $i < j$*,* π *is* less precise *than* π' *if* $i > j$*, and* π *is* as precise as π' *if* $i = j$*.*

Since the set of predictions in Π_1 play a crucial role in what is to follow, it is worth explaining exactly why they have this degree of precision. Essentially it is because the form of qualitative probabilistic inference on which $\mathcal{QPR}/\mathcal{QPR}_{\mathcal{C}}$ is based is relativistic—it computes changes without taking account of the previous value. Thus when we obtain a prediction of $\Uparrow$ it is given with no guarantee that the probability, before whatever change led to this prediction, was not already 1. So it is not possible to guarantee that a change will actually occur. Something similar is true of $\uparrow$, though here, even if the initial value was known, there would still be imprecision in the prediction because of the $\geq$ (or $\leq$) in the definition of the probabilistic constraints which give rise to such predictions.[3]

In addition to defining the relative precision of two predictions, we can define what it means if they agree. Intuitively, two predictions agree if they predict changes that can be reconciled. This if one prediction is that the probability of p is 0, and another prediction is that the probability of p will increase to 1, then the two disagree. Alternatively, a prediction of $\uparrow$ and another of $\Uparrow$ agree since both predict changes in the same direction and the latter is just more precise about the state that results from the prediction.

To formally define what agreement is, we need to distinguish between predictions of increases and predictions of decreases:

Definition 8 *The set of change predictions* Π_1 *is the union of* $\Pi_1^{\uparrow} = \{\Uparrow, \uparrow\}$*, the* increasing *predictions, and* ${\Pi_1}^{\downarrow} = \{\Downarrow, \downarrow\}$*, the* decreasing *predictions.*

Note that $\Pi_1^{\uparrow} \cap {\Pi_1}^{\downarrow} = \emptyset$. We also have:

Definition 9 *The set of* categorical *predictions* Π^{cat} *is the union of* $\Pi_0^{cat} = \{1, 0\}$*, the* categorical value *predictions, and* $\Pi_1^{cat} = \{\Uparrow, \Downarrow\}$*, the* categorical change *predictions.*

We call the categorical value predictions the *limit* predictions, and thus 1 is known as the *upper* limit prediction, and 0 as the emphlower limit prediction.

Definition 10 *The set of* non-categorical *predictions* $\Pi^{non-cat}$ *is the union of* $\Pi_0^{non-cat} = \{\wr\}$*, the* non-categorial value *prediction,* $\Pi_1^{non-cat} = \{\uparrow, \downarrow\}$*, the* non-categorial change *predictions, and the set of* vacuous *predictions* $\Pi_2 = \{\updownarrow\}$*.*

[3] Although it is possible to define systems which do give precise predictions about changes in value, systems that produce precise predictions are more awkward to work with than those that don't [11].

Note that $\Pi^{cat_c} \cap \Pi^{non-cat} = \emptyset$.

At this point we should recall the discussion about the flattening function following Definition 4. The existence of the blank spaces in Table 2 can be explained in terms of the following theorem which shows that it is not possible to have categorical predictions for a proposition which conflict:

Theorem 11 *It is impossible to have:*

1. *an increasing and decreasing categorical prediction; or*
2. *an increasing categorical prediction and a lower limit prediction; or*
3. *a decreasing categorical prediction and an upper limit prediction; or*
4. *an upper and lower limit prediction*

for the same proposition.

Proof: We prove this by contradiction. Assume we have two arguments for p, one with an increasing categorical prediction and one with a decreasing categorical prediction. For this to be the case, there must be one of the following pairs of categorical implications (since only categorial implications can generate categorical predictions): $\langle (i : q \to p : ++), (j : r \to p : --) \rangle$, $\langle (i : q \to p : ++), (i : r \to p : +-) \rangle$, $\langle (i : q \to p : --), (i : r \to p : -+) \rangle$, $\langle (i : q \to p : +-), (i : r \to p : -+) \rangle$, or symmetrical variations. Now, consider the constraints on the joint probability distribution over p, q and r imposed by the first pair. By definition, these imply that $\Pr(p|q, X) = 1$ and $\Pr(p|r, Y) = 0$ for all X and Y, which would require $\Pr(p|q, r)$ to be both 1 and 0. Thus the first pair of implications cannot occur together. Similarly the constraints embodied by the second pair would require $\Pr(p|q, \neg r)$ to be 1 and 0, the third pair would also require $\Pr(p|q, \neg r)$ to be 1 and 0 and the fourth pair would require $\Pr(p|\neg q, \neg r)$ to be 1 and 0. Since these are impossible, increasing and decreasing categorical predictions cannot occur together. The remaining parts of the result are proved similarly. □

This theorem gives us our first taste of dealing with several predictions for the same formula. Since a single prediction is related to a single argument, and we typically have several arguments for a given formula, we often need to deal with several predictions at a time. In fact, given some proposition p, what we are interested in is the overall prediction after flattening all the arguments in the argument set for p. Thus we have:

Definition 12 *The* overall prediction *of the argument set* $\mathbf{A}_p^{\triangle}$ *is* π_O *where* $\mathsf{flat}(\mathbf{A}_p^{\triangle}) = \langle p, \pi_O \rangle$.

If all the arguments in an argument set $\mathbf{A}_p^{\triangle}$ make non-categorical predictions, then the argument set is said to be *non-categorical*. Otherwise it is said to be *categorical*. The reason for this distinction is because of the nature of the overall predictions made by these sets:

Theorem 13 *The overall prediction of an argument set* $\mathbf{A}_p^{\triangle}$ *is non-categorical if and only if* $\mathbf{A}_p^{\triangle}$ *is non-categorical. The overall prediction of an argument set* $\mathbf{A}_p^{\triangle}$ *is categorical if and only if* $\mathbf{A}_p^{\triangle}$ *is categorical.*

Proof: The "if" part of the theorem is as follows: If $\mathbf{A}_p^{\triangle}$ is non-categorical, then, by definition, all predictions made by arguments in $\mathbf{A}_p^{\triangle}$ are non-categorical. From the definition of $\oplus$, flattening these predictions will give a non-categorical prediction. Conversely, if $\mathbf{A}_p^{\triangle}$ is categorical then there is at least one argument in $\mathbf{A}_p^{\triangle}$ which makes a categorical prediction. From Theorem 11 we know that we can't have conflicting categorical predictions for any proposition, and from the definition of $\oplus$ we know that combining a categorical and non-categorical prediction will give a categorical prediction. Thus, it follows that from the definition of $\oplus$ flattening a mixed set of categorical and non-categorical predictions will give a categorical prediction.

The "only if" part is as follows. If the overall prediction is non-categorical then, from the definition of $\oplus$, the argument set cannot contain any categorical predictions. Thus the argument set is non-categorical. If the overall prediction is categorical then, again from the definition of $\oplus$, there must be at least one categorical prediction in the argument set and so the argument set is categorical. $\square$

There are a couple of important corollaries of this result:

Corollary 14 *Adding a categorical argument to a non-categorical argument set gives a categorical argument set.*

Corollary 15 *Adding a non-categorical argument to a categorical argument set gives a categorical argument set.*

Now, the intuitive notion of agreement between predictions introduced above is partly based on precision. A no change prediction agrees with an increasing prediction (since an increasing prediction is a prediction of "either an increase, or no change") or a decreasing prediction, and for similar reasons, an increasing or decreasing prediction will agree with a vacuous prediction.

In fact, any two-non-categorical predictions will agree, since it is always possible to find some probability, or change in probability, that reconciles the two predictions. Even predictions of $\uparrow$ and $\downarrow$ can be reconciled if the actual change being predicted is 0. Thus at least one of two predictions that disagree has to be categorical, and a little thought shows that the particular categorical values of such predictions that cannot be reconciled are those which Theorem 11 rules impossible. In addition an increasing prediction will disagree with a categorical decrease (but not a lower limit prediction since the probability might start at 0 and not change).

Thus we have: Thus we have:

Definition 16 *Two predictions* π *and* π' disagree *iff:*

1. *one is an increasing prediction and the other is a decreasing categorical prediction; or*
2. *one is an decreasing prediction and the other is a increasing categorical prediction; or*
3. *one is an increasing categorical prediction and the other is a lower limit prediction; or*
4. *one is a decreasing categorical prediction and the other is an upper limit prediction; or*
5. *one is an upper limit prediction and the other is a lower limit prediction*

Two predictions are said to *agree* if they do not disagree.

This makes the idea of disagreement rather weak, but it is the only sensible notion—given the tentative notion of an increase that we are dealing with, this notion of disagreement is the only one that can be related to unsoundness with respect to probability theory. That is, if two predictions disagree according to Definition 16 and one is sound with respect to probability theory (in other words it is not possible to rule out an increase in probability), then the other is incorrect. If the second prediction comes from either Π_0 or Π_2 then it will not be unsound with respect to probability theory since there is a change in probability, consistent with all that is known, which can satisfy both predictions.

Now, the reason for this long digression from the solution of the problem of not dealing with d-separation, is to be able to prove the following results:

Lemma 17 *Given two non-categorical argument sets $\mathbf{A}_p^{\Delta}$ and $\mathbf{A}_p^{\Delta'}$ about a proposition p, with overall predictions π_O and π'_O, such that $\mathbf{A}_p^{\Delta} \subset \mathbf{A}_p^{\Delta'}$, π'_O will agree with π_O.*

Proof: The only time that π_O and π'_O can disagree is if one is increasing and the other is decreasing. Consider π_O to be increasing. Adding additional non-categorical predictions to π_O using $\oplus$ will give an overall prediction which is either increasing or vacuous. Thus the new overall prediction will not disagree with π_O, though it may be less precise. If π_O is decreasing, the same argument can be applied. Thus the two predictions cannot disagree, and the result follows. $\Box$

When argument sets are non-categorical, the predictions of individual arguments are closely related to the signs propagated along trails in a QPN [2], and the overall prediction of the argument set $\mathbf{A}_p^{\Delta}$ which contains all the arguments for p which can be built from a database is closely related to the sign which is produced by the sign-propagation algorithm for QPNs [3]. As a result, Lemma 17 can be taken as a form of robustness result for inference in QPNs.

It is worth noting that although it might be tempting to read it that way, Lemma 17 does not mean that qualitative probabilistic inference is somehow immune to unsoundness (and therefore inherently uninteresting). What the lemma says is that provided there is at least one non-categorical prediction which agrees with probability theory, adding in further non-categorical predictions will not make the overall prediction disagree with the original. (All that will happen is that as both increasing and decreasing predictions are added the overall prediction will become vacuous.) This behaviour is therefore a reflection of the robustness of the flattening procedure more than anything else.

We can also consider the case where we have categorical arguments:

Lemma 18 *Given two categorical argument sets $\mathbf{A}_p^{\Delta}$ and $\mathbf{A}_p^{\Delta'}$ about a proposition p, with overall predictions π_O and π'_O, such that $\mathbf{A}_p^{\Delta} \subset \mathbf{A}_p^{\Delta'}$, π'_O will agree with π_O.*

Proof: From Theorem 13 both π_O and π'_O will be categorical, and from Theorem 11 they cannot disagree, so they must agree. $\Box$

In addition to these two cases, there is one in which the first argument set is non-categorical and the second is categorical:

Lemma 19 *Given a non-categorical argument set $\mathbf{A}_p^{\Delta}$ and a categorical argument set $\mathbf{A}_p^{\Delta'}$ about a proposition p, with overall predictions π_O and π'_O, such that $\mathbf{A}_p^{\Delta} \subset \mathbf{A}_p^{\Delta'}$, π'_O may disagree with π_O.*

Proof: Here all we need to show is that it is possible for π_O and π'_O to disagree, and we can do that by example. From Theorem 13 π_O will be non-categorical while by Corollary 14 π'_O will be categorical. Consider that π_O is increasing and π'_O is decreasing (because π'_O is the result of adding an argument with sign $\Downarrow$ to $\mathbf{A}_p^{\Delta}$), then π'_O will disagree with π_O. If the argument that is added has sign $\Uparrow$, then the two predictions will agree. $\Box$

The final case that one might imagine, where $\mathbf{A}_p^{\Delta}$ is categorical and $\mathbf{A}_p^{\Delta'}$ is non-categorical, is ruled out by Corollary 15 since adding arguments to a categorical argument set will give a categorical argument set. Taking all these results together, we can conclude that adding a non-categorical argument to an argument set will never cause the new overall prediction to disagree with the old overall prediction.

Now we can finally identify when, in general, the predictions of two argument sets can disagree, thus homing in on when the problem with causal inference in $\mathcal{QPR}/\mathcal{QPR}_{\mathcal{C}}$, as currently defined, will be significant. This is when a categorical argument is added to an argument set.

Theorem 20 *Given two argument sets $\mathbf{A}_p^{\Delta}$ and $\mathbf{A}_p^{\Delta'}$ about a proposition p, with overall predictions π_O and π'_O, such that $\mathbf{A}_p^{\Delta} \subset \mathbf{A}_p^{\Delta'}$, π'_O will only disagree with π_O when:*

1. *$\mathbf{A}_p^{\Delta}$ is non-categorical and $\mathbf{A}_p^{\Delta'}$ is categorical; and*
2. *either π_O is increasing and and π'_O is decreasing or a lower limit, or π_O is decreasing and π'_O is increasing or an upper limit*

Proof: For the first part, consider $\mathbf{A}_p^{\Delta'}$ to be formed by adding arguments to $\mathbf{A}_p^{\Delta}$. Since, by Corollary 15, adding arguments to a categorical argument set will give a categorical argument set, there are only three possibilities for the two sets: (i) they are both categorical, in which case by Lemma 18 it is the case that π_O and π'_O agree; (ii) they are both non-categorical, in which case by Lemma 17 it is the case that π_O and π'_O agree, or (ii) $\mathbf{A}_p^{\Delta'}$ is non-categorical and $\mathbf{A}_p^{\Delta'}$ is categorical when, by Lemma 19 and the proof thereof, it is the case that π_O and π'_O will disagree if one is increasing and the other is decreasing. The second part follows directly from the definition of disagreement. □

This result concerns general argument sets. The specific case that we are interested in is when the extra arguments in $\mathbf{A}_p^{\Delta'}$, over and above those in $\mathbf{A}_p^{\Delta}$, should be ruled out by d-separation. In the next section we formally define arguments that should be ruled out by d-separation and denote them as *invalid*. Borrowing that terminology without proper definition for now, we can state the case we are interested in, that for which there is a problem, as being that when $\mathbf{A}_p^{\Delta}$ contains all valid arguments for p (all those that are not invalid), and $\mathbf{A}_p^{\Delta'}$ contains all the valid arguments and some invalid ones as well.

In such a case we can see from Theorem 20 that the problem will arise only when the invalid arguments added to $\mathbf{A}_p^{\Delta}$ include at least one categorical argument and where either the overall prediction of $\mathbf{A}_p^{\Delta}$ is increasing and the overall prediction of $\mathbf{A}_p^{\Delta'}$ is decreasing or the overall prediction of $\mathbf{A}_p^{\Delta}$ is decreasing and the overall prediction of $\mathbf{A}_p^{\Delta'}$ is increasing. Since by Lemmas 17–19 only adding a categorical prediction can cause disagreement, the problem arises when one of the invalid arguments is categorical and in the opposite direction to the overall prediction of the valid ones.

As an aside, it should be noted that since categorical arguments are rather rare (since they require every implication chained together as part of the argument to be not only categorical but also categorical in the right direction)[4] it seems unlikely that ignoring d-separation will cause major problems. However, the problem does need to be fixed, and this is the subject of the next section.

5 Correct causal inference

The discussion so far has identified where the flaw in $\mathcal{QPR}/\mathcal{QPR}_\mathcal{C}$ lies. It occurs because additional arguments, over and above those which should be flattened, are combined in the flattening process—ones that should be excluded by d-separation—and these overturn the prediction that should be made.

[4]Bearing in mind that combining some categorical implications with some categorical predictions does not yield a categorical prediction, see Table 1.

5.1 Bringing in *d*-separation

How should we take d-separation into account when flattening causal arguments?[5] One way is to use the same technique that $\mathcal{QPR}$, as described in [10], used when dealing with combined causal and evidential reasoning and combined causal, evidential and intercausal reasoning. This technique needs the following definitions:

Definition 21 *In the triple* $(i : l : s)$*, the* wff l *is said to be* indexed *by the symbol* i*.*

Definition 22 *A* source *of an argument* (p, G, s) *is an* swff *indexed by an element of* G*.*

Thus a source of an argument is one of the simple formula which ground it, and form the head of a chain of implications.

Definition 23 *The* destination *of an argument* (p, G, s) *is* p*.*

Thus the destination of an argument is the formula being argued for.

Definition 24 *Two formulae* p *and* q *are* d-separated *if* p *or* q *has probability 1 or 0, or if for all arguments which have* p *as a source and* q *as their destination, there is another formula* r *such that either:*

1. *p is a cause of r, r is a cause of q, and the probability of r is 1 or 0; or*
2. *r is a cause of p, r is a cause of q and the probability of r is 1 or 0; or*
3. *p and q are both causes of r and there is no argument (r, G', s') such that all the* swffs *indexed by elements of G' are effects of r, and the probability of r is not 1 or 0.*

The first item defines the form of d-separation missing from $\mathcal{QPR}_{\mathcal{C}}$. The second item is required in evidential reasoning, and the third is required in intercausal reasoning. Despite the fact that we are only concerned with causal reasoning in $\mathcal{QPR}_{\mathcal{C}}$, we include all these forms of d-separation in order to have a definition that will work in every situation. An invalid argument is now one that is built without taking account of d-separation:

Definition 25 *An argument* $A = (p, G, s)$ *is* invalid *if any source of* A *is* d*-separated from* p*.*

[5] One might argue that a better solution is to take account of d-separation when constructing arguments, but since identifying some forms of d-separation requires all arguments to be constructed, as is clear from the definition of d-separation given below, it seems conceptually simpler to simply build all arguments and rule out the ones which don't take account of d-separation.

This notion of invalidity differs from that introduced in [10] only because the notion of d-separation which underpins invalidity has been expanded. Thus this new notion of invalidity rules out more arguments from flattening and hence from having an effect on the overall prediction of an argument set.

Now, we have:

Definition 26 *An argument $A = (p, G, s)$ is* valid *if it is not invalid.*

Now that we can identify which arguments are valid, and hence can be helpfully flattened, all that is necessary is to redefine the procedure for determining the change in probability of some formula p. The new procedure is:

1. Add a triple $(i : q : s)$ for every formula q whose change in probability is known.
2. Build $\mathbf{A}_p^{\Delta}$ using the C-rules.
3. Flatten this set to $\mathsf{Flat}'_{\mathsf{c}}(\mathbf{A}_p^{\Delta})$ where $\mathsf{Flat}'_{\mathsf{c}}(\mathbf{A}_p^{\Delta}) = \mathsf{flat}(\mathsf{flat}_{\mathsf{c}}(\mathbf{A}_p^{\Delta}))$.

where:

$$\mathsf{flat}_{\mathsf{c}} : \mathbf{A} \mapsto \{A \in \mathbf{A} \mid A \text{ is valid}\}$$

With this change to the flattening function, $\mathcal{QPR}/\mathcal{QPR}_{\mathcal{C}}$ will not generate spurious results by ignoring d-separation and so will be sound.

5.2 Back to $\mathcal{QPR}$

What we have found, therefore, is that in order to correct the inference carried out in $\mathcal{QPR}_{\mathcal{C}}$, we have to check arguments before flattening, and only flatten those arguments which are valid in the sense defined above. These are arguments which have been built using the set of C-rules and are not ruled out by d-separation. The question we are interested in here is how, then, can we take this result and use it to modify the original system $\mathcal{QPR}$ in order to make that correct as well. To do this we need to recall how the original system was defined.

As discussed above, $\mathcal{QPR}$ was defined in much the same way as $\mathcal{QPR}_{\mathcal{C}}$ is here, but rather than having just one set of proof rules it has three. $\mathcal{QPR}$ has the same set of C-rules as were presented here plus a set of E-rules which permit evidential reasoning, and a set of I-rules which permit intercausal reasoning. Separate proof procedures were given for reasoning with the C-rules alone (it is this "mode of inference" of the overall system[6] we have been discussing here), with the C-rules and the E-rules, and with the C-rules, E-rules and I-rules. As initially defined, the procedure for using the C-rules was just that given earlier in the paper—build arguments and then flatten them with flat; the procedure for the C-rules and E-rules in combination was to build arguments, rule out those made ineligible by d-separation or by using the same implication more

[6]The scare quotes are used since these different forms of inference were really an artifact of the fact that I chose to construct the soundness and completeness proofs incrementally rather than because I expected $\mathcal{QPR}$ to be used in three distinct ways.

than once (which rules out cyclic arguments) and flatten those; when using all three sets of proof rules in conjunction we use the same procedure as for the C-rules and E-rules.

With this description and the discussion in Section 5.1, it is clear how to patch $\mathcal{QPR}$. When just using the C-rules, we apply $\mathsf{flat_c}$ to the set of valid arguments; when using the C-rules and E-rules we remove invalid arguments and non-minimal arguments (the name given in [10] for cyclic arguments), and we adopt the same procedure when using all three sets of rules. This neatly glosses over another problem with [10], which is that the definition of d-separation given there is itself flawed—it is missing the second clause in Definition 24—which would lead to the same kind of problems when using the C-rules and E-rules together as we have investigated here for the C-rules. Of course, using the revised version of $\mathcal{QPR}$ described here will correct both problems. This revised version of $\mathcal{QPR}$ is described in full in [12], a modified version of [10].

5.3 The issue of many pieces of evidence

One interesting fact follows from the new flattening function. In [10] it was claimed that when used for causal reasoning, the construction of arguments was entirely local, that is once an argument was built, the change it predicted could not be ruled out by further inference. Clearly this is not true of the revised version of $\mathcal{QPR}$ since flattening will now rule out arguments whose sources are d-separated from their destinations. Now, to some extent this lack of locality is an implementation issue. After all, in our example it is possible to identify that a and c are d-separated while the argument is being constructed. However, the issue is more complex—consider what would happen in Δ_2 or Δ_3 if instead of knowing that b was true, we knew that d was true when we also had $(i : d \rightarrow b : {+}{+})$ in our database. In that case the probability of b would still be 1, but in order to know this, it would first be necessary to construct all arguments for b and flatten them. Thus the order in which arguments are constructed becomes important, and that opens up a whole new set of issues.

To illustrate these issues, consider the following example:

Example 5. The following clauses represent the fact that a has a positive influence on b, and b has a positive influence on c while d has a negative influence on b.

$$\begin{array}{ll} (r1 : a \rightarrow b : +) & \Delta_5 \\ (r2 : b \rightarrow c : +) & \\ (r3 : d \rightarrow b : -) & \end{array}$$

Now, consider we have evidence that $\Pr(a)$ and $\Pr(d)$ are observed to be true, so that the triples $(f1 : a : \Uparrow)$ and $(f1 : d : \Uparrow)$ are added to the database, it is possible to build two arguments concerning $\Pr(c)$ using $\mathcal{QPR}_{\mathcal{C}}$:

$$\begin{array}{lll} \Delta_5 & \vdash_{QP} & (c, \{f1, r1, r2\}, \uparrow) \\ \Delta_5 & \vdash_{QP} & (c, \{f2, r3, r2\}, \downarrow) \end{array}$$

The first is built by combining $f1$ and $r1$ using →-E, and then using the result of this inference with $r2$ using →-E again. The second is built by combining $f1$ and $r2$ using →-E and then chaining the result of this with $r3$ using →-E again. These two arguments may then be flattened to give the pair $\langle c, \ \updownarrow\rangle$. □

This is entirely correct. Because we have no information about the strength of the influences, there is no way to resolve the tradeoff between the positive influence on a on c and the negative influence on d on c (though see [9, 15, 17] for approaches to resolving such tradeoffs). Now consider a small variation on this new example.

Example 6. Here d has a categorical influence on b:

$$\begin{array}{ll} (r1 : a \rightarrow b : +) & \quad \Delta_6 \\ (r2 : b \rightarrow c : +) & \\ (r3 : d \rightarrow b : --) & \end{array}$$

Again consider we have evidence that $\Pr(a)$ and $\Pr(d)$ are observed to be true, so that the triples $(f1 : a : \Uparrow)$ and $(f1 : d : \Uparrow)$ are added to the database. We get the same two arguments about $\Pr(c)$ using $\mathcal{QPR_C}$:

$$\begin{array}{lll} \Delta_6 & \vdash_{QP} & (c, \{f1, r1, r2\}, \ \uparrow) \\ \Delta_6 & \vdash_{QP} & (c, \{f2, r3, r2\}, \ \downarrow) \end{array}$$

As ever, they may then be flattened to give the pair $\langle c, \ \updownarrow\rangle$. □

This third example is again not entirely correct, and here the problem is quite subtle, and points to issues at the heart of qualitative probabilistic reasoning.

The nut of the problem is this. Qualitative probabilistic reasoning as captured in QPNs and $\mathcal{QPR}/\mathcal{QPR_C}$ is concerned with changes in probability in response to evidence. As such the results of qualitative probabilistic inference make perfect sense when a single piece of evidence is presented. All the changes calculated are those that result from that single piece of evidence. Under this "single evidence assumption" algorithms for propagation in QPNs and associated models [2, 3, 11], including $\mathcal{QPR}/\mathcal{QPR_C}$, are sound. The problem arises when there are several pieces of evidence, and this matter is further explored, in the context of qualitative probabilistic networks, in [16]. To some extent, this problem of multiple pieces of evidence is at the heart of the flaws in $\mathcal{QPR}/\mathcal{QPR_C}$ that we have been concerned with here—as pointed out above, $\mathcal{QPR}$ works fine so long as there is no old evidence. Future work will further investigate this problem.

6 Summary

This paper has identified a problem with the system of qualitative probabilistic argumentation introduced in [10]. The paper precisely defined the problem and then proceeded to explore its effects and then establish a solution. The problem with the original system was the fact that it failed to take d-separation into account in causal reasoning, a failing that is easy to correct by ruling out arguments which do not respect d-separation (exactly the solution provided here).

References

[1] S. Benferhat, D. Dubois, and H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, 1993.

[2] M. J. Druzdzel. *Probabilistic reasoning in decision support systems: from computation to common sense*. PhD thesis, Carnegie Mellon University, 1993.

[3] M. J. Druzdzel and M. Henrion. Efficient propagation in qualitative probabilistic networks. In *Proceedings of the 11th National Conference on Artificial Intelligence*, 1993.

[4] P. M. Dung. On the acceptability of arguments and its fundamental role in non-monotonic reasoning and logic programming. In *Proceedings of the 13th International Conference on Artificial Intelligence*, 1993.

[5] J. Fox. A unified framework for hypothetical and practical reasoning (2): lessons from clinical medicine. In *Proceedings of the Conference on Formal and Applied Practical Reasoning*, 1996.

[6] F. V. Jensen. *An introduction to Bayesian networks*. UCL Press, London, UK, 1996.

[7] B. Kuipers, editor. *Qualitative Reasoning—Modelling and Simulation with Incomplete Knowledge*. MIT Press, Cambridge, MA, 1994.

[8] R. Loui. Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, 3:100–106, 1987.

[9] S. Parsons. Refining reasoning in qualitative probabilistic networks. In *Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence*, 1995.

[10] S. Parsons. A proof theoretic approach to qualitative probabilistic reasoning. *International Journal of Approximate Reasoning*, 19:265–297, 1998.

[11] S. Parsons. *Qualitative approaches to reasoning under uncertainty.* MIT Press, Cambridge, MA, 2001.

[12] S. Parsons. A proof theoretic approach to qualitative probabilistic reasoning—revised version. Technical report, Department of Computer and Information Science, Brooklyn College, City University of New York, 2002.

[13] S. Parsons and S. Green. Argumentation and qualitative decision making. In A. Hunter and S. Parsons, editors, *Symbolic and Quantitative Approaches to Reasoning and Uncertainty*, pages 328–339. Springer Verlag, Berlin, Germany, 1999.

[14] J. L. Pollock. Justification and defeat. *Artificial Intelligence*, 67:377–407, 1994.

[15] S. Renooij and L. van der Gaag. Enhancing QPNs for trade-off resolution. In K. B. Laskey and H. Prade, editors, *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, pages 559–566, San Francisco, CA, 1999. Morgan Kaufmann.

[16] S. Renooij, L. van der Gaag, and S. Parsons. Propagation of multiple observations in QPNs revisited. In *Proceedings of the 15th European Conference on Artificial Intelligence*, 2002.

[17] S. Renooij, L.C. van der Gaag, S. Parsons, and S. Green. Pivotal pruning of trade-offs in QPNs. In C. Boutillier and M. Goldszmidt, editors, *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA, 2000. Morgan Kaufmann.

[18] M. P. Wellman. *Formulation of tradeoffs in planning under uncertainty*. Pitman, London, 1990.

[19] M. P. Wellman and M. Henrion. Qualitative intercausal relations, or explaining "explaining away". In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 535–546, San Mateo, CA, 1991. Morgan Kaufmann.