Arguing about potential causal relations

Leila Amgoud Henri Prade

IRIT, Université de Toulouse, CNRS

Reasoning about causality

i) **Deductive causal reasoning**

- generic causal relations
- particular situation

⇒ predict what is going to take place

(generally) A causes B
A is true

B should be true (and might be expected to be reported as such)

ii) **Abductive reasoning**

- generic causal relations
- observed facts

⇒ diagnose plausible causes

(generally) A causes B
B is true

A might be true

iii) **causality assessment**

- reported sequences of facts
- generic knowledge about the (normal) course of the world

⇒ identify the causal relation(s) between the reported facts

iv) **analogical reasoning**

- past experience:

set of reported sequences of facts with identified causal relations

⇒ guess causal relations in a new reported sequence of facts
(on a similarity basis)

v) **inductive reasoning**

- a *sufficiently large* **set** of reported sequences

⇒ learn generic causal relations.

Case (iii)

context C B_t, A_t, $\neg B_{t'}$ $t' > t$

Definitions

- a sequence B_t, A_t, $\neg B_{t'}$ is reported to an agent
- agent's knowledge: nonmonotonic consequence relation $\mid\!\approx$

Facilitation C : $A_t \Rightarrow$fa $\neg B_{t'}$

if C $\mid\!\approx$ B and C $\wedge$ A $\not\mid\!\approx$ B

A_t is perceived as having facilitated the occurrence of $\neg B_t$, in context C

Causation C : $A_t \Rightarrow$ca $\neg B_{t'}$

if C $\mid\!\approx$ B and C $\wedge$ A $\mid\!\approx \neg$B

A_t is perceived as being the cause of $\neg B_t$, in context C

• If C: A ⇒ca B, or if C: A ⇒fa B, then C |≈ ¬ A

• ***restricted*** **transitivity**

If C: A ⇒ca B, if C: B ⇒ca D and if **B ∧ C |≈ A**

then C: A ⇒ca D

holds for ⇒ca if |≈ is a preferential entailment

holds for ⇒fa if |≈ is a rational closure entailment.

B ∧ C |≈ A

the normal way to have B (in context C) is to have A

A = drinking, B = inebriated, D: staggering,

'drinking' ⇒ca 'inebriated' 'inebriated' ⇒ca staggering'

'drinking' ⇒ca 'staggering'

'inebriated' |≈ 'drinking'

Justification (or Explanation)

sequence: B_t, A_t, $\neg B_{t'}$

Agent's knowledge:

$C \not\mid\approx B$, $C \not\mid\approx \neg B$ and $C \wedge A \mid\approx \neg B$

($\mid\approx$ non-monotonic consequence relation)

A is perceived as *justifying / explaining* the fact that B is now false in context C

Different possible scenarios

- $C, B_t, \neg B_{t'}$ change without reported event
- $C, B_t, B_{t'}$ persistence without reported event
- $C, B_t, A_t, \neg B_{t'}$ change with reported event
- $C, B_t, A_t, B_{t'}$ persistence with reported event

possible pieces of knowledge

- either $C \mid\!\approx B$, or $C \mid\!\approx \neg B$, or $C \not\mid\!\approx B$ and $C \not\mid\!\approx \neg B$
- either $C \wedge A \mid\!\approx B$, or $C \wedge A \mid\!\approx \neg B$, or

$C \wedge A \not\mid\!\approx B$ and $C \wedge A \not\mid\!\approx \neg B$

36 scenarii

1	C, Bt, At, ¬Bt'	C $\mid\!\approx$ B	C∧A $\mid\!\approx$ B	unexplained change, B should have persisted
2	C, Bt, At, ¬Bt'	C $\mid\!\approx$ B	C∧A $\not\mid\!\approx$ B and C∧A $\not\mid\!\approx$ ¬B	change *facilitated* by A
3	C, Bt, At, ¬Bt'	C $\mid\!\approx$ B	C∧A $\mid\!\approx$ ¬B	change *caused* by A
4	C, Bt, At, ¬Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ ¬B	C∧A $\mid\!\approx$ B	unjustified change after A
5	C, Bt, At, ¬Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ ¬B	C∧A $\not\mid\!\approx$ B and C∧A $\not\mid\!\approx$ ¬B	contingent change
6	C, Bt, At, ¬Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ ¬B	C∧A $\mid\!\approx$ ¬B	change *justified* by A
7	C, Bt, At, ¬Bt'	C $\mid\!\approx$ ¬B	C∧A $\mid\!\approx$ B	unexplained change, double defeated expectations!
8	C, Bt, At, ¬Bt'	C $\mid\!\approx$ ¬B	C∧A $\not\mid\!\approx$ B and C∧A $\not\mid\!\approx$ ¬B	from exceptionality to contingency
9	C, Bt, At, ¬Bt'	C $\mid\!\approx$ ¬B	C∧A $\mid\!\approx$ ¬B	back to normality thanks to A

10	C, Bt, ¬Bt'	C $\mid\!\approx$ B	C∧A $\mid\!\approx$ B	change for unknown reason
11	C, Bt, ¬Bt'	C $\mid\!\approx$ B	C∧A $\not\mid\!\approx$ B and C∧A $\not\mid\!\approx$ ¬B	change for unknown reason, A is a potential facilitating factor
12	C, Bt, ¬Bt'	C $\mid\!\approx$ B	C∧A $\mid\!\approx$ ¬B	A is a potential *cause* for the change
13	C, Bt, ¬Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ ¬B	C∧A $\mid\!\approx$ B	unexplainable change
14	C, Bt, ¬Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ ¬B	C∧A $\not\mid\!\approx$ B and C∧A $\not\mid\!\approx$ ¬B	fully contingent change
15	C, Bt, ¬Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ ¬B	C∧A $\mid\!\approx$ ¬B	A would *justify* the change
16	C, Bt, ¬Bt'	C $\mid\!\approx$ ¬B	C∧A $\mid\!\approx$ B	back to normality (not due to A)
17	C, Bt, ¬Bt'	C $\mid\!\approx$ ¬B	C∧A $\not\mid\!\approx$ B and C∧A $\not\mid\!\approx$ ¬B	back to normality, (could have been facilitated by A)
18	C, Bt, ¬Bt'	C $\mid\!\approx$ ¬B	C∧A $\mid\!\approx$ ¬B	back to normality (maybe due to A)

19	C, Bt, At, Bt'	C $\mid\!\approx$ B	C$\wedge$A $\mid\!\approx$ B	A agrees with persistence of B
20	C, Bt, At, Bt'	C $\mid\!\approx$ B	C$\wedge$A $\not\mid\!\approx$ B and C$\wedge$A $\not\mid\!\approx$ $\neg$B	B has persisted in spite of A
21	C, Bt, At, Bt'	C $\mid\!\approx$ B	C$\wedge$A $\mid\!\approx$ $\neg$B	unexplained persistence of B
22	C, Bt, At, Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ $\neg$B	C$\wedge$A $\mid\!\approx$ B	A explains persistence of B
23	C, Bt, At, Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ $\neg$B	C$\wedge$A $\not\mid\!\approx$ B and C$\wedge$A $\not\mid\!\approx$ $\neg$B	contingent persistence of B
24	C, Bt, At, Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ $\neg$B	C$\wedge$A $\mid\!\approx$ $\neg$B	A disagrees with persistence of B
25	C, Bt, At, Bt'	C $\mid\!\approx$ $\neg$B	C$\wedge$A $\mid\!\approx$ B	back to normality
26	C, Bt, At, Bt'	C $\mid\!\approx$ $\neg$B	C$\wedge$A $\not\mid\!\approx$ B and C$\wedge$A $\not\mid\!\approx$ $\neg$B	from exception to contingency
27	C, Bt, At, Bt'	C $\mid\!\approx$ $\neg$B	C$\wedge$A $\mid\!\approx$ $\neg$B	double defeated expectations, exceptional situation

28	C, Bt, Bt'	C $\mid\!\approx$ B	C$\wedge$A $\mid\!\approx$ B	expected persistence
29	C, Bt, Bt'	C $\mid\!\approx$ B	C$\wedge$A $\not\mid\!\approx$ B and C$\wedge$A $\not\mid\!\approx$ $\neg$B	expected persistence
30	C, Bt, Bt'	C $\mid\!\approx$ B	C$\wedge$A $\mid\!\approx$ $\neg$B	expected persistence
31	C, Bt, Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ $\neg$B	C$\wedge$A $\mid\!\approx$ B	contingent persistence
32	C, Bt, Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ $\neg$B	C$\wedge$A $\not\mid\!\approx$ B and C$\wedge$A $\not\mid\!\approx$ $\neg$B	contingent persistence
33	C, Bt, Bt'	C $\not\mid\!\approx$ B and C $\not\mid\!\approx$ $\neg$B	C$\wedge$A $\mid\!\approx$ $\neg$B	contingent persistence
34	C, Bt, Bt'	C $\mid\!\approx$ $\neg$B	C$\wedge$A $\mid\!\approx$ B	from exception to normality in case A took place
35	C, Bt, Bt'	C $\mid\!\approx$ $\neg$B	C$\wedge$A $\not\mid\!\approx$ B and C$\wedge$A $\not\mid\!\approx$ $\neg$B	persistence of exceptionality, might be facilitated to A
36	C, Bt, Bt'	C $\mid\!\approx$ $\neg$B	C$\wedge$A $\mid\!\approx$ $\neg$B	persistence of exceptionality

Arguing causality

- Argument : a reason for claiming that event A causes B

  $\Rightarrow$ A causes B is not necessarily true

  $\Rightarrow$ argument may be attacked by other arguments

- Argumentation : reasoning about interacting arguments

Causal Argument

Definition: Causal argument scheme

A is an *arguable cause for* ¬ B because:

a. Normally in context C, B is true $C \mid\!\approx B$

b. The actual context is $C' = C \wedge A$

(assuming consistency of C and A)

c. In the new context C', ¬B is reported as true

A *relevant or (significant) difference*

between contexts C and C'

Example

A bicyclist moves into the traffic lane in order to pass a truck illegally parked in the bike lane. The driver of a car approaching from the rear slams on her brakes in order to avoid hitting the bicycle. A following car fails to stop in time, and smashes into the back of the first.

The bicyclist's insurance company may claim that the illegally parked truck (i) caused her client to swerve (s) into the lane of traffic, using

Argument A: i caused s because:

a. $C \mid\!\approx \neg s$

b. $C' = C \wedge i$

c. s is true

Critical questions

- Does it hold that C $|\!\approx$ B ? Are there cases where C $\wedge$ $\neg$B holds?
- Is it really the case that $\neg$B is true?
- Is there another A' such that both C $\wedge$ A' and $\neg$B hold?
- Is the difference A pointed out between contexts C and C' relevant

  (w. r. t. a possible change from B to $\neg$B)?
- Does the possible cause A invariably, or at least generally, produce the effect $\neg$B?

$\Rightarrow$ answering the above questions amounts to exhibit counter-arguments

$\Rightarrow$ one or several of the prototypical situations listed in Table

Example

several persons get sick after eating a pizza during a party organized by their friend Mary. Moreover, each of them had a fancy hat also.

Argument A_1: *pizza* caused *sick* because:

party $|\approx \neg$ sick

C' = party $\wedge$ pizza

sick is true.

Argument A_2: *wearing a hat* caused *sick* because:

party $|\approx \neg$ sick

C' = party $\wedge$ wearing a hat

sick is true.

Argument A_3: wearing a hat $|\approx \neg$ sick

not a causal argument!

fancy hats were treated by means of some toxic product

Argument A_4: toxic product caused sick because:

wearing a hat $|\approx \neg$ sick

C' = wearing a hat $\wedge$ toxic product

sick is true

argumentation is a dynamical process where arguments and counter-arguments interact with each other in order to assess a possible cause

Concluding remarks

- to figure out what may be the different types of reaction an agent may have in face of a sequence, depending on his beliefs on the normal course of things
- causal arguments, where do they come from, and how they may be refuted.
- Dung's acceptability semantics are not suitable in case of causal arguments

looking for **responsibility**

"If A' had taken place, ¬B would not have happened"

- A' is an uncontrolled event

"if no storm had taken place, there would be no flood"

- A' is an action performed by some agent

"if Peter had abstained drinking, he would not have got a fee"

Here ¬B is something **undesirable**

A' may be regarded as a cause for it

But similar patterns exist where ¬B is **desirable**

"if Peter had not received a solid education, he would have not succeeded"

"if embankments had not be built, the flood would have not been avoided"

condition part may appear either in a positive or negative form