Iterated Abduction

Joshua Eckroth

Stetson University

FLAIRS-29

Outline

- Review of *abduction*.
- Demonstration of the problem of *iterated abduction*.
- Prior work.
- Implementation of iterated abduction.
- Examples.
- Conclusion.

The famously wet grass

- The grass is wet. Why?
- Perhaps the sprinkler was on.
- Or perhaps it rained.

Abduction is the process of finding an explanation for the wet grass (viz., sprinkler or rain).

Abduction

- Logic-based abduction [e.g., Aliseda, 2006]: find consistent *abducibles* that, if asserted, would entail or support a proof of the evidence.
- Set-covering abduction [e.g., Josephson and Josephson, 1994]: find a consistent set of explanations that cover as much of the evidence as possible.
- Probabilistic (Bayesian) abduction [e.g., Pearl, 1988]: having observed E, find C such that $P(E|C)$ is maximized.

Iterated abduction

- The grass is wet.
- That's because it was raining!
- Oh wait, the weather report says it was not raining.
- Ok, it was the sprinkler then!

Given w, $r \to w$, and $s \to w$, abduce r. Now, suppose $\neg r$. Thus, abduce s (automatically).

Iterated abduction

Suppose,

- Rain causes the grass to be wet. $r \to w$
- Sprinklers cause the grass to be wet. $s \to w$
- The sprinklers are never on when it is raining. $\neg s \lor \neg r$
- Only rain can make the neighbor's grass wet. $r \to w'$

Upon learning our grass is wet (w), we may abduce that the sprinklers were on (s). Upon learning our neighbor's grass is wet (w'), we further abduce rain (r); however, rain is inconsistent with sprinklers, so we contract our belief that the sprinklers were on, leaving rain to explain both wet grasses.

Desideratum for iterated abduction

- The epistemic rules ensure that alternative explanations for prior evidence are abduced whenever supporting explanations are contracted.

We want a system that supports expansion, contraction, abduction, and iterated forms of these.

Prior work

- In most prior work [e.g., Aliseda, 2006], each abduction is considered novel rather than building off prior abductions. Thus, the desideratum is not met.
- Nayak and Foo [1999] developed a method of iterated abduction that left open alternative explanations as long as possible; the worst explanation was eliminated at each step.
- Eckroth and Josephson [2014] developed a metareasoning system that attempted to identify which prior explanations should be contracted; but it was complex and lacking formal specification.
- Beirlaen and Aliseda [2014] described a conditional logic for abduction that identifies the explanations that are still viable after other alternative explanations have been defeated; their work only handles belief expansion.

Computational complexity

- Finding minimal abductions is NP-complete [Bylander et al., 1991].
- Finding minimal contractions is NP-complete [Tennant, 2012].
- Thus, any practical system must employ heuristics for finding good enough abductions and contractions.

Finite dependency networks

For the purposes of developing algorithms for minimal and near-minimal contractions, Tennant [Tennant, 2012] introduced the *finite dependency network.*

> *Belief systems can be represented as finite dependency networks. Structureless nodes represent the agent's beliefs. Nodes are either 'in' or 'out' (believed or not believed) . [Strokes] connecting the nodes represent 'apodeictic' warrant-preserving transitions (for the agent). [Strokes] are either initial or transitional.*

Tennant used the colors *black* to represent believed propositions and *white* to represent disbelieved propositions.

Example FDN

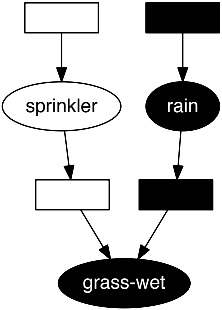


Figure 1: The wet grass example as an FDN.

Example FDN

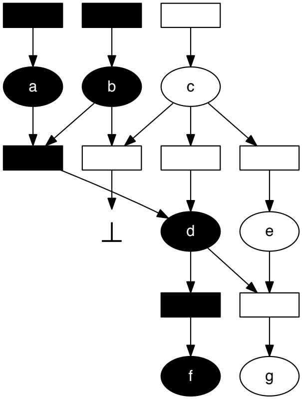


Figure 2: The equivalent belief system is:
$\{\{a, b\} \rightsquigarrow d, c \rightsquigarrow d, c \rightsquigarrow e, d \rightsquigarrow f, \{d, e\} \rightsquigarrow g, \{b, c\} \rightsquigarrow \perp,$
$a, b, \neg c, d, \neg e, f, \neg g, \neg\perp\}$, where $\rightsquigarrow$ means "can explain."

Axioms of coloration

- (C1) Every black node receives an arrow from some black inference stroke, i.e., $\forall x((Bx \wedge Nx) \rightarrow \exists y(By \wedge Sy \wedge Ayx))$.
- (C2) Every white node receives arrows (if any) only from white inference strokes, i.e., $\forall x((Wx \wedge Nx) \rightarrow \forall y(Ayx \rightarrow (Wy \wedge Sy)))$.
- (C3) Every black inference stroke receives arrows (if any) only from black nodes, i.e., $\forall x((Bx \wedge Sx) \rightarrow \forall y(Ayx \rightarrow (By \wedge Ny)))$.
- (C4) Every white inference stroke that receives an arrow receives an arrow from some white node, i.e., $\forall x((Wx \wedge Sx \wedge \exists zAzx) \rightarrow \exists y(Wy \wedge Ny \wedge Ayx))$.
- (C5) The node $\perp$ is white, i.e., $W\perp$.

Axioms of coloration: C1

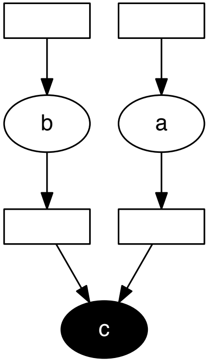


Figure 3: Violation of axiom C1: "Every black node receives an arrow from some black inference stroke."

Axioms of coloration: C2

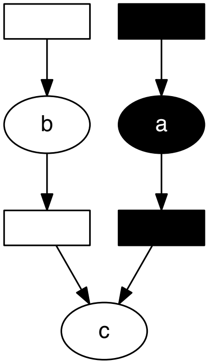


Figure 4: Violation of axiom C2: "Every white node receives arrows (if any) only from white inference strokes."

Axioms of coloration: C3

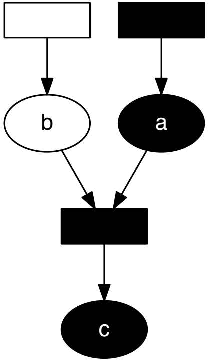


Figure 5: Violation of axiom C3: "Every black inference stroke receives arrows (if any) only from black nodes."

Axioms of coloration: C4

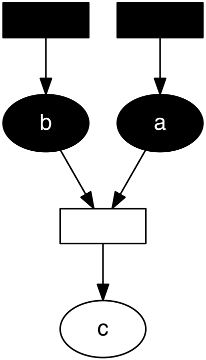


Figure 6: Violation of axiom C4: “Every white inference stroke that receives an arrow receives an arrow from some white node.”

Axioms of coloration: C5

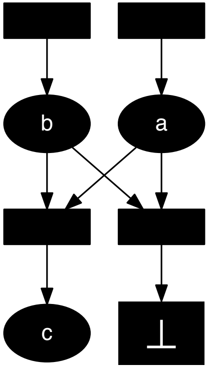
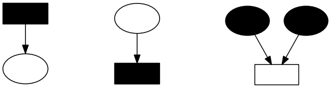
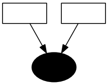


Figure 7: Violation of axiom C5: “The node $\perp$ is white.”

Abduction: Bad strokes and nodes

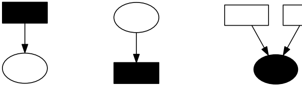


Deterministic bad strokes and nodes in abduction. In each case, the white node or stroke should be black.

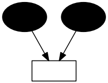


Nondeterministic bad strokes in abduction. At least one white stroke should be black.

Contraction: Bad strokes and nodes



Deterministic bad strokes and nodes in contraction. In each case, the black node or stroke should be white.



Nondeterministic bad nodes in contraction. At least one black node should be white.

Abduction and contraction in FDNs

- The patterns of invalid coloration for abduction and contraction are sufficient to define an algorithm that restores consistency after abducing (turning black) or contracting (turning white) a particular node.
- The algorithm could simply identify all bad nodes and strokes, alter their color, and then repeat until the axioms of coloration are met.
- But this algorithm is not capable of finding alternative explanations for previously explained evidence when it contracts those explanations.

Grass example (broken)

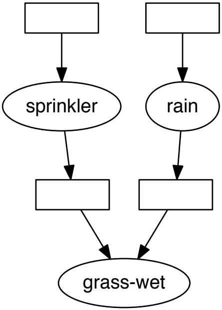


Figure 8: Initially, no beliefs are held.

Grass example (broken)

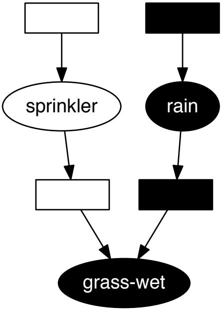


Figure 9: Upon learning that the grass is wet, rain is abduced.

Grass example (broken)

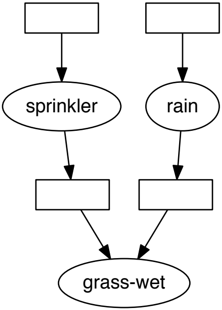


Figure 10: Upon learning that rain is false, the belief about wet grass is contracted as well.

Solution: Add priorities

- Iterated abduction requires that at the agent know the order of abductions and contractions, specifically that grass-wet was abduced before rain was contracted, and that sprinkler was not (recently) abduced or contracted.
- This order of node and stroke color changes means that sprinkler remains a possible explanation for grass-wet but rain does not (after rain is contracted).

Solution: Add priorities

- We can record the priority (or time) of color changes by defining a function $T(\cdot)$ that maps nodes and strokes to the set of natural numbers. $T(x) = t$ means that node or a stroke x acquired its current color at time t.
- Whenever the system acquires a new observation, the global time counter is incremented.
- Each nodes and stroke retains its priority value until it changes color again.

Grass example with priorities

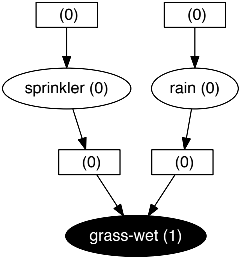


Figure 11: Time 1, observed grass-wet. Consistency-restoration has not yet occurred.

Grass example with priorities

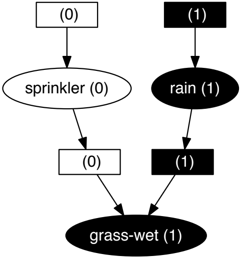


Figure 12: Suppose the stroke from rain is chosen to turn black. The rain node and its incoming stroke are likewise turned black. The axioms of coloration are now satisfied.

Consistency-restoration with priorities

Bad nodes and strokes that are black, respecting priority

$$
\begin{aligned}
\mathcal{B}_{\mathcal{SD}} &= \{s|Ss \land Bs \land \\
&\qquad ([\exists n : Asn \land Wn \land T(s) \leq T(n)] \lor \\
&\qquad\ [\exists n' : An's \land Wn' \land T(s) \leq T(n')])\} \\
\mathcal{B}_{\mathcal{ND}} &= \{n|Nn \land Bn \land (\forall s : Asn \to Ws) \land \\
&\qquad (\forall s : Asn \to T(n) \leq T(s))\} \\
\mathcal{B}_{\mathcal{NN}} &= \{n|Nn \land Bn \land \\
&\qquad (\exists s : Ans \land Ws \land \\
&\qquad\ (\forall n' : An's \to Bn') \land T(n) \leq T(s))\}
\end{aligned}
$$

Consistency-restoration with priorities

Bad nodes and strokes that are white, respecting priority

$$
\begin{aligned}
\mathcal{W}_{\mathcal{SD}} &= \{s | Ss \wedge Ws \wedge (\forall n : Ans \rightarrow Bn) \wedge \\
&\qquad (\exists n : Ans \wedge T(s) < T(n))\} \\
\mathcal{W}_{\mathcal{SN}} &= \{s | Ss \wedge Ws \wedge \\
&\qquad (\exists n : Asn \wedge Bn \wedge \\
&\qquad\ (\forall s' : As'n \rightarrow Ws') \wedge T(s) < T(n))\} \\
\mathcal{W}_{\mathcal{ND}} &= \{n | Nn \wedge Wn \wedge \\
&\qquad ([\exists s : Asn \wedge Bs \wedge T(n) < T(s)] \vee \\
&\qquad\ [\exists s : Ans \wedge Bs \wedge T(n) < T(s)])\}
\end{aligned}
$$

Consistency-restoration algorithm

- If the FDN satisfies the axioms of coloration, we are done.
- Otherwise, let $\mathcal{D} = \mathcal{B}_{\mathcal{SD}} \cup \mathcal{B}_{\mathcal{ND}} \cup \mathcal{W}_{\mathcal{SD}} \cup \mathcal{W}_{\mathcal{ND}}$, i.e., all deterministic bad nodes and strokes. Change the color of all of these (black to white or white to black), and repeat at step (1).
- If no deterministic bad nodes or strokes exist, then let $\mathcal{N} = \mathcal{B}_{\mathcal{NN}} \cup \mathcal{W}_{\mathcal{SN}}$, i.e., all nondeterministic bad nodes and strokes. Select one of these according to a heuristic and change its color. Repeat at step (1).

Grass example with priorities

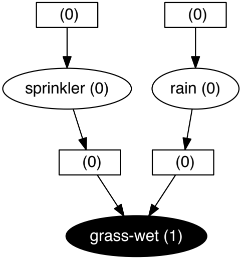


Figure 13: Time 1, observed grass-wet. Both strokes pointing to grass-wet are bad strokes (in $\mathcal{W}_{\mathcal{SN}}$), so a choice must be made.

Grass example with priorities

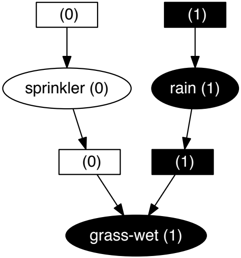


Figure 14: Suppose the stroke from rain is chosen to turn black. The rain node and its incoming stroke are likewise turned black. The axioms of coloration are now satisfied.

Grass example with priorities

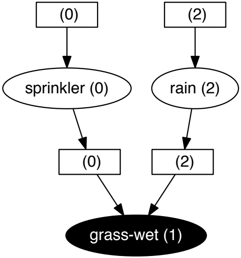


Figure 15: Time 2, contracted rain; whitening spreads to nearby strokes. The grass-wet node does not meet the criteria for a bad node, since an incoming stroke has lower priority than grass-wet.

Grass example with priorities

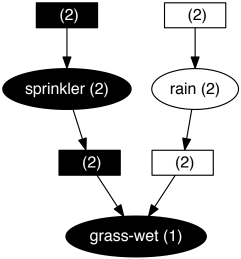


Figure 16: The only bad stroke or node is the stroke coming from sprinkler (the single member of $\mathcal{W}_{\mathcal{SN}}$), so it is turned black. Ultimately, consistency is restored again.

Lung cancer example with priorities

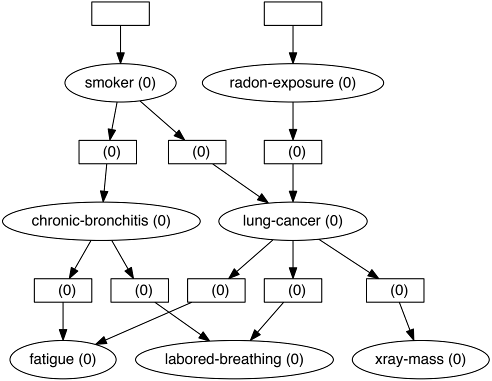


Figure 17: Initial state of the lung cancer model.

Lung cancer example with priorities

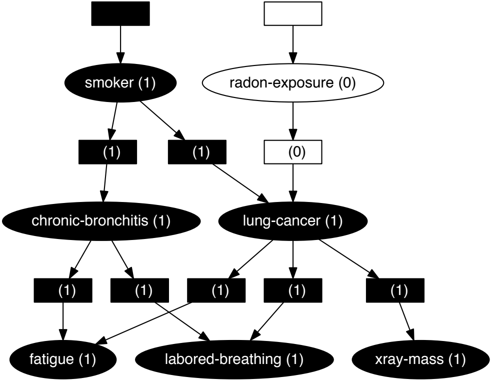


Figure 18: Upon observing fatigue and labored breathing, chronic bronchitis and smoker are abduced, which result in beliefs lung cancer and xray mass.

Lung cancer example with priorities

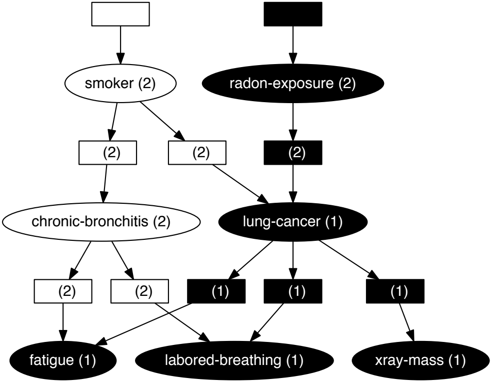


Figure 19: Then, upon learning the patient is not a smoker, radon exposure is abduced instead.

Lung cancer example with priorities

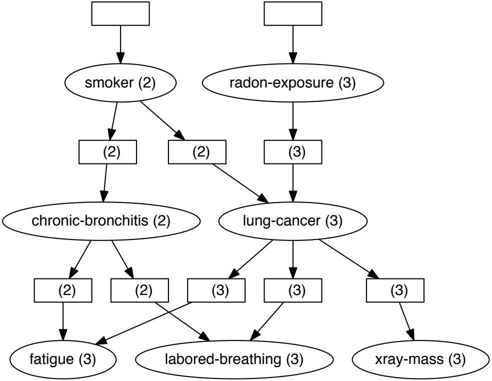


Figure 20: Finally, suppose the patient also reports no radon exposure. Given the closed world assumption, there is now no explanation for fatigue and labored breathing, so they are contracted.

Conclusion

- We characterized the problem of *iterated abduction* and proposed a desideratum for epistemic rules.
- We explored a computational encoding of the epistemic rules with Tennant's *finite dependency networks*.
- By adding priorities to FDNs and modifying the consistency-restoration algorithm, we met the desideratum of iterated abduction with a computational implementation.

Future work

- Add predicates and variables, e.g., *parent(x,y)* and *male(x)* can explain *father(x,y)*.
- Evaluate contraction and abduction heuristics: minimality, explanatory coverage, recovery.
- Evaluate benchmark cases, e.g., plan recognition, story understanding, et al.
- Evaluate scalability, e.g., ABox abduction and the "semantic web."

Predicates and variables

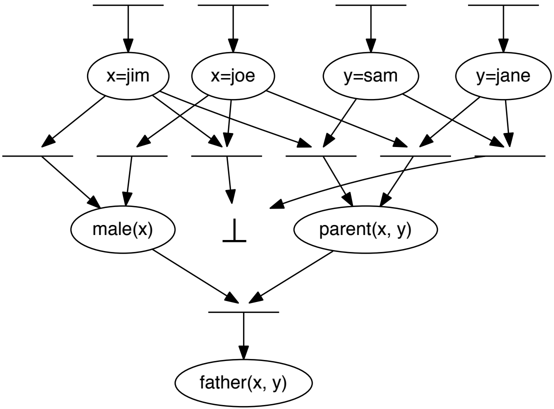


Figure 21: father(x,y), parent(x,y), and male(x) predicates and possible instantiations.

Predicates and variables

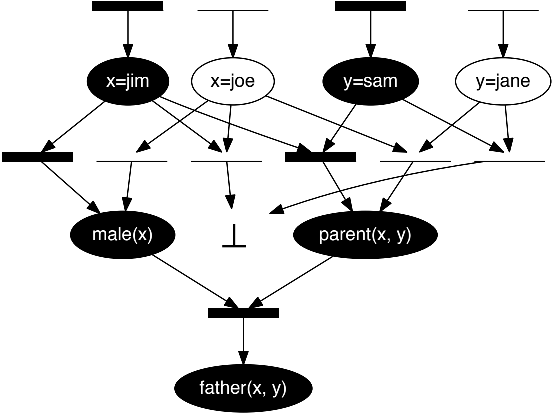


Figure 22: Result from abduction after learning father(x,y).

Predicates and variables

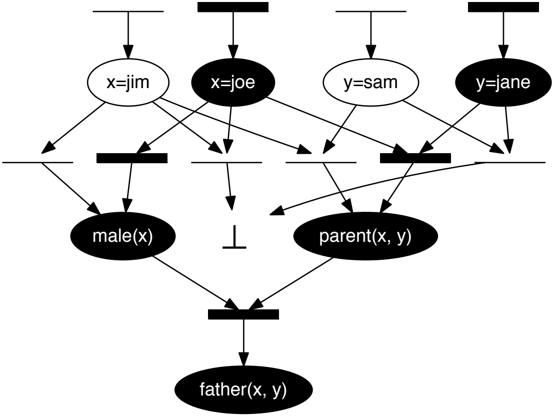


Figure 23: Result after subsequently contracting x=jim.

Predicates and variables

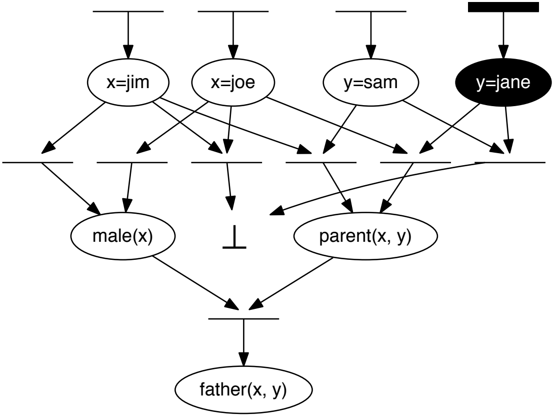


Figure 24: Result after subsequently contracting x=jim and x=joe.

Questions?

Formal properties

- Let $\mathcal{L}$ be a propositional language and $p \in \mathcal{L}$ identify a statement in that language.
- Let $\rightsquigarrow \subset \mathcal{P}(\mathcal{L}) \times (\mathcal{L} \cup \{\perp\})$ be a relation between statements such that $P \rightsquigarrow q$, where $P = \{p_1, \ldots, p_n\} \subset \mathcal{L}$, means "the conjunction $\cap P$ can, if true, explain q."
- We write $p \rightsquigarrow q$ as a notational shorthand for $\{p\} \rightsquigarrow q$.
- A statement q may be *assumable*, i.e., requiring no explanation, which we denote as $\mapsto q$ as a shorthand for $\emptyset \rightsquigarrow q$.

Formal properties

- A set P of inconsistent statements may be denoted $P \rightsquigarrow \bot$; in this case, not all statements in P may be consistently simultaneously believed by the agent.
- We assume the set of explanatory relations believed by the agent is exhaustive ("closed-world assumption").
- Let $\mathbb{B}$ be the set of explanatory relations (including assumables) and other beliefs that the agent holds at any time.

Belief system criteria

$\mathbb{B}$ is a *belief system* if it meets the following criteria:

- $\forall p :\mapsto p \in \mathbb{B} \to A(p)$.
- $\forall P, q : P \rightsquigarrow q \in \mathbb{B} \to (\forall p \in P : A(p)) \wedge A(q)$.
- $\forall p : A(p) \to (p \in \mathbb{B} \vee \neg p \in \mathbb{B})$.
- $\forall P, q : P \rightsquigarrow q \in \mathbb{B} \to q \notin P$, i.e., atoms cannot directly participate in explaining themselves.
- $\forall P, q : P \rightsquigarrow q \in \mathbb{B} \wedge P \subset \mathbb{B} \to q \in \mathbb{B}$, i.e., explanations imply what they explain.
- $\neg\bot \in \mathbb{B}$, thus ensuring $\mathbb{B}$ is internally consistent. It is assumed, but not written for brevity's sake, that $\{p, \neg p\} \rightsquigarrow \bot$ for all atomic statements p.
- $\forall q : (A(q) \wedge (q \in \mathbb{B})) \to (\mapsto q \in \mathbb{B} \vee (\exists P : P \rightsquigarrow q \in \mathbb{B} \wedge (P \subset \mathbb{B})))$, i.e., every belief that requires explanation is explained by some other belief(s).

Contraction and abduction

- *Contraction* is incorporating $\neg q$ into $\mathbb{B}$ for some $q \in \mathcal{L}$, denoted $\mathbb{B} - q$.
- *Abduction* is incorporating q into $\mathbb{B}$ for some $q \in \mathcal{L}$, denoted $\mathbb{B} + q$.

The criteria for belief systems ensure that consistency is maintained and every belief that requires an explanation indeed has an explanation.

Belief system / FDN isomorphism

A belief system is isomorphic to an FDN in the following way.

- Atomic statements are nodes in the FDN:
  $(A(q) \wedge (\mapsto\ q \in \mathbb{B} \vee (\exists P : P \rightsquigarrow q \in \mathbb{B})) \leftrightarrow Nq.$
- Believed atomic statements are black nodes in the FDN:
  $(A(q) \wedge q \in \mathbb{B}) \leftrightarrow Bq.$
- The $\rightsquigarrow$ relation is represented by a stroke in the FDN:
  $P \rightsquigarrow q \in \mathbb{B} \leftrightarrow (\exists s : Ss \wedge (\forall p \in P : Aps) \wedge Asq).$
- An assumable atomic statement p is represented by a stroke with no incoming arrows:
  $\mapsto\ p \in \mathbb{B} \leftrightarrow (\exists s : Ss \wedge Asp \wedge \neg(\exists p' : Ap's)).$

References I

A. Aliseda. *Abductive reasoning: Logical investigations into discovery and explanation*. Kluwer Academic Publishers, Norwell, MA, 2006.

Mathieu Beirlaen and Atocha Aliseda. A conditional logic for abduction. *Synthese*, 191(15):3733–3758, 2014.

T. Bylander, D. Allemang, M.C. Tanner, and J.R. Josephson. The computational complexity of abduction. *Artificial Intelligence*, 49 (1-3):25–60, 1991.

Joshua Eckroth and John R. Josephson. Anomaly-driven belief revision and noise detection by abductive metareasoning. *Advances in Cognitive Systems*, 3:123–143, 2014.

J. R. Josephson and S. G. Josephson. *Abductive inference: Computation, philosophy, technology*. Cambridge University Press, Cambridge, UK, 1994.

References II

Abhaya C Nayak and Norman Y Foo. Abduction without minimality. In *Advanced Topics in Artificial Intelligence*, pages 365–377. Springer, 1999.

J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

Neil Tennant. *Changes of Mind: An Essay on Rational Belief Revision*. Oxford University Press, 2012.