Memory, Explanation and Creativity: SWALE

David Leake, Indiana University

1

Some Historical Roots from Research on Story Understanding

- Much early CBR grew out of Yale story understanding research
- Their first attempts at story understanding viewed understanding as being able to build up causal chains of inferences to establish coherence

David Leake, Indiana University

2

2

Building Up Connections by Causal Chaining: A Simplified Example

3

Pros and Cons of Causal Chaining

- Pro: Causal chains enable inferring related steps and answering questions
- Cons:
  - Chaining can be expensive
  - Chains may not reflect context

4

4

Building Up Connections: Another Example

John was hungry.

Hungry -> Goal to eat -> Needs food -> Go to source

He went to the refrigerator.

Want something + Item available -> Ask

He asked for a hamburger.

5

Scripts (Schank & Abelson, 77)

- Scripts record standard events in stereotyped contexts (e.g., restaurant, laundry, airport,…)
- The restaurant script might include *enter, be seated, order, receive food, eat, pay, leave…*
- Once an appropriate script is selected, that script provides context-relevant inferences

6

Pros and Cons of Scripts

- Pros:
  - Efficiency
  - Identifying important inferences (shared background for summarization, etc.)
  - Cognitive validity
- Cons:
  - Lack of flexibility
  - Problems as a model of memory: Script confusions

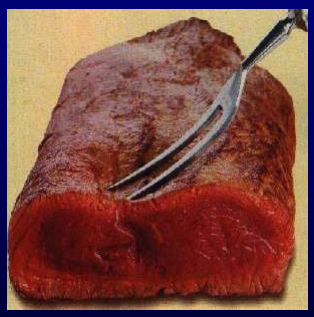
Beyond Scripts

- When scripts fail can give important clues to human processing
- Anomalous events can prompt *remindings* during understanding
- The study of reminding is the study of how memories are organized, retrieved, and applied
- This is central to case-based reasoning

Dynamic Memory Theory

- Schank's *Dynamic memory* (1982), develops roots of CBR by analyzing remindings.

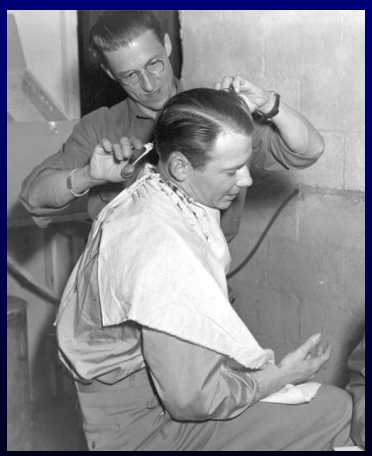
"X described to Y how his wife would never cook his steak as rare as he liked it."



How is this episode indexed? We can tell by the remindings it prompts.

What Comes to Mind from the Steak Story?

"On hearing the steak story, Y was reminded of a time, 30 years earlier, when he tried to get his hair cut in England and the barber just wouldn't cut it as short as he wanted it."



Dynamic Memory Theory tries to account for this.

Dynamic Memory Theory, part 1

- Dynamic Memory Theory tries to account for
  - Cross-contextual reminding and learning
  - Observed human memory confusions
- Episodes are stored under processing structures called Memory Organization Packages (MOPs)
- MOPs are hierarchical and interconnected

A MOP for Professional Office Visits

This provides:
- Property inheritance
- Reconstructive memory, so…
- Memory confusions in people

How Are Remindings Indexed Under MOPs?

- A good index must support being reminded at the right time
- A good index must support *not* being reminded at the wrong time

How Would You Index the Steak and Haircut?

- Schank's answer:
  - *Provide-service* – the knowledge structure governing processing when the failure occurred
  - *Refusal because request too extreme* – the explanation.

Dynamic Memory Theory (part 2)

- Failures are stored under the structures that failed, indexed by the explanations
- Explanations correspond to abstract themes, called Thematic Organization Points (TOPs)
- Track your own remindings to see if this holds for you!

How Can Remindings be Useful?

- Breakout groups: When a failure occurs and you're reminded of a similar failure, how can the reminding be useful?

How Can Remindings be Useful?

- Schank: Failure-driven remindings can help for
  - Scoping
  - Verification (of an explanation)
  - Response information
- How far can the use of remindings go?

CBR View: Remindings Permeate Reasoning

Does thinking really involve thinking?

(Riesbeck & Schank, 1989)

Case Study: Modeling CBR in Understanding in SWALE
(Schank & Leake 1989; Leake 92; Schank, Riesbeck, Kass, 1994)

- SWALE models the role of context in indexing, retrieval, and solution evaluation for explanation of anomalous events
- The system's namesake example is explaining the story of the racehorse Swale:

Swale was a star 3-year-old racehorse, winning all the most important races.

A few days after a major victory, he returned from a light morning gallop and collapsed, dead, at his stable.

19

Why Explaining Swale is Hard for AI Systems

- Limited information about circumstances
- Computational cost of chaining through possible causes
- Imperfect and incomplete knowledge
- Note: Explainable AI is currently one of the hottest AI areas

20

However, People Can Do It

Why did Swale die?

Swale's Death Prompted Remindings

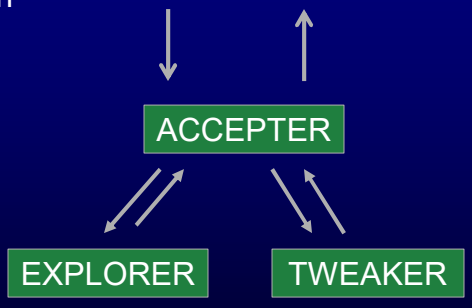
- A vet: "This sounds like an aneurysm. I've seen this sort of thing before."
- A Yale AI lab student: "This sounds like the death of Jim Fixx " (a runner, who died of a heart-attack when in peak condition.)
- Another lab student: "Swale was a young superstar like Janis Joplin. Maybe he died of a drug overdose." Could this apply to Swale?

Motivations for a Case-Based Approach

- Addressing classic problems in story understanding
  - The combinatorial explosion of inference chaining
  - The lack of flexibility of schema-based understanding
- Modeling a creative process

SWALE's Explanation Process

- Anomaly detection
- Anomaly characterization
- Explanation retrieval
- Explanation evaluation
- Explanation adaptation



Adaptation of Remindings is Crucial

- A major focus was on achieving flexible reuse
- Different remindings have different levels of applicability
  - A vet: "This sounds like an aneurysm. I've seen this sort of thing before."
  - A Yale AI lab student: "This sounds like the death of Jim Fixx " (a runner, who died of a heart-attack when in peak condition.)
  - Another lab student: "Swale was a young superstar like Janis Joplin. Maybe he died of a drug overdose."

SWALE's Adaptation (Kass, 1992)

- The SWALE system aimed at developing a flexible "middle way" between specific and general adaptation knowledge.
- The project also tried to find general ways to access and apply adaptation rules
- One focus was developing a taxonomy of possible problems requiring adaptation.

SWALE's Adaptation (Kass, 1992)

- The SWALE system aimed at developing a flexible "middle way" between specific and general adaptation knowledge.
- The project also tried to find general ways to access and apply adaptation rules
- One focus was developing a taxonomy of possible problems requiring adaptation.

How Should Cases be Represented?

- Recall that representations are shaped by tasks
- Here, the representation should support *retrieval*, *adaptation*, and *evaluation*
- Breakout groups: Sketch a representation for a case for Swale's death, with the explanation that Swale died when from the exertion of racing and a heart defect.

SWALE's Case Representation

```
mop isa (m-xp)
 observed: m-healthy-animal
 expected: m-unhealthy-animal
 events: ((jog . m-horserace-event)
          (exert . m-exert-event)
          (defect . m-heart-defect-state)
          (outcome . m-dead-state))
 causals: ((jog supports exert)
           (defect exert supports outcome))
 constraints: (((events jog actor)
                (events exert actor)
                (events defect owner)
                (events outcome object)))
```


SWALE's Problem Types

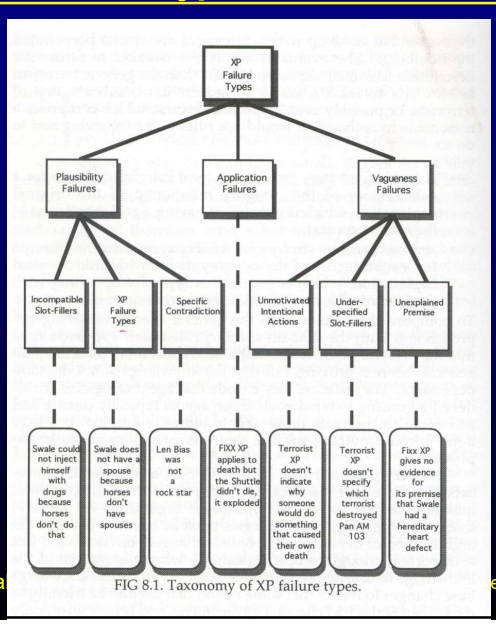


FIG 8.1. Taxonomy of XP failure types.

Tweaking Strategies

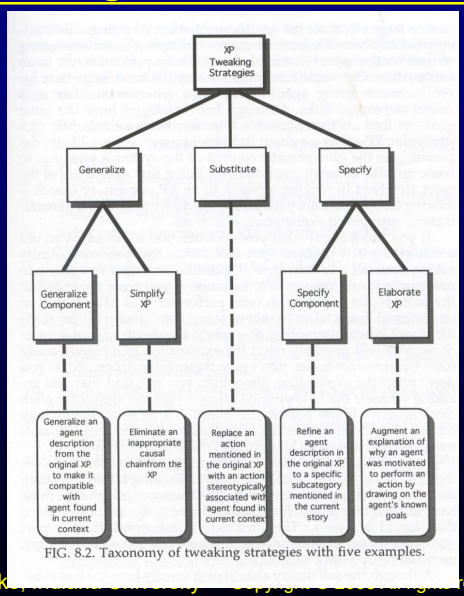


FIG. 8.2. Taxonomy of tweaking strategies with five examples.

32

Levels of Case Adaptation

- Different levels of adaptation provide different levels of flexibility
- Directly applicable explanation: An aneurysm
  - Applicability problems: None
  - Adaptation: Substitute Swale for previous victim
  - Result: Swale died of an aneurysm
- When no adaptation Is required, CBR is basically re-using stored solutions

33

- More distinct explanation: Heart attack from jogging
  - Problem: Horses aren't joggers
  - Adaptation: Search memory for something racehorses do with the same effects
  - Adaptation strategy: To find replacement actions, look for theme actions
  - Result: Find that racehorses have exertion during racing and training runs
  - Result: Swale died of a heart attack during a training run
- Routine differences require straightforward repair; adaptation generates a "similar" solution

How Jogging is Adapted

- Receives an explanation and problem description as input.
- Retrieves adaptation strategies associated with the problem type.
- The strategies are rules for finding the information needed to revise explanations and for revising them.
- Applies the rules to generate a new explanation.

Adapting to Fix "Horses Can't Jog"

- Input problem description is ACTION-AGENT-MISMATCH
- System retrieves adaptation strategies including:
  - *Replace action: Use agent-theme links*
- Searches memory for role themes of Swale.
- Finds ``running in races" and ``eating oats"
- Attempts to substitute.
- Running in races is a form of physical exertion and fits the chain.
- Result: Swale's exertion running in races overtaxed a hereditary heart defect.

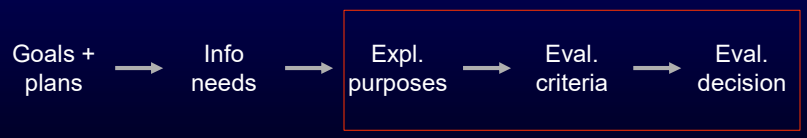
36

- Janis Joplin explanation:
  - Stress ⇨ recreational drug use ⇨ overdose
- Problems: Racehorses can't take recreational drugs.
- Adaptation: Remove recreational drugs, keep drug overdose, and try to justify. Result:
  - *Swale died when his trainer gave an overdose of performance-enhancing drugs.*
- Lesson: When differences are major, adaptation can result in radically different solutions

37

Evaluation is Part of CBR's Reasoning too: Explanations Are Evaluated by Purpose

- ACCEPTER selects evaluation criteria suitable to high-level *explanation purposes* such as prediction, repair, or assigning responsibility.
- Each explanation purpose is associated with evaluation dimensions such as predictive power, timeliness, routineness, distinctiveness, knowability, causal force, repairability, independence, achievability/blockability, and desireability.
- Example: Useful prediction requires predictive power, knowability, timeliness, and distinctiveness



Transcript of a MicroSWALE Run

```
> (explain-swale)

Micro Explainer
===============
Trying to explain m-swale-story which has the anomaly:
  instance isa (m-event-sequence-conflict)
  event-sequence: m-racehorse-life
  expectation: m-expectation-1
  surprising-fact: m-swale-dead-state

Micro XP Retriever
==================
Searching for XP's.
... found:
    m-die-from-illness-xp, m-die-from-old-age-xp.
Applying m-die-from-old-age-xp.
... generated m-expl1.
Applying m-die-from-illness-xp.
... generated m-expl2.
```


Micro-ACCEPTER – Evaluation of Illness XP

```
Micro-Accepter
==============
Evaluating m-expl2 for the purpose of prediction.
The underling XP is:
  mop isa (m-xp)
  events: ((illness . m-bad-health-state) (outcome . m-dead-state))
  causals: ((illness => outcome))
  constraints: (((events illness object) (events outcome object)))
Checking if explanation accounts for m-swale-dead-state.
... appears relevant.
Checking believability.
Problem found!
  instance isa (m-non-normative-filler-problem)
  explanation: m-expl2
  path: (events illness object)
  expected-value: m-unhealthy-animal
  observed-value: m-swale
... explanation is not believable.
```


Micro-Accepter (continued)

```
Checking if explanation is useful for the purpose of prediction.
Is the explanation predictive?
... Explanation is adequate for prediction.
Problem found: m-prob3
Checking seriousness of problems.
... m-expl2 is tweakable
```


Micro Tweaker – Tweaking Illness XP

Micro Tweaker
==========
Looking for candidate tweaks in memory for m-prob3.
... found:
m-replace-action-use-stereotypes-tweak,
m-replace-actor-use-stereotypes-tweak,
m-replace-action-use-causals-tweak.
Trying tweak replace-action-search-through-stereotypes on m-expl2.
Looking for effects to be accounted for.
... found m-dead-state
Looking for scripts for m-swale that include m-dead-state
... no scripts found.
... tweak failed.
Trying tweak replace-actor-search-through-stereotypes on m-expl2.
[tweak unimplemented]
... tweak failed.
Trying tweak replace-action-search-through-causal-rules on m-expl2.
[tweak unimplemented]
... tweak failed.

Micro XP Retriever – Seeking XPs under "death while healthy"

Micro XP Retriever
==========
Searching for XP's.
Looking for retrieval cues in m-prob3.
... Cue: observed m-swale, rather than m-unhealthy-animal.
... found m-fixx-xp
Applying m-fixx-xp.
... generated m-expl4.

Micro Accepter – Checking Old Age XP

Micro-Accepter
==========
Evaluating m-expl1 for the purpose of prediction.
The underling XP is:
  mop isa (m-xp)
  events: ((old-age . m-old-age-state) (outcome . m-dead-state))
  causals: ((old-age => outcome))
  constraints: (((events old-age object) (events outcome object)))
Checking if explanation accounts for m-swale-dead-state.
... appears relevant.

Checking believability.
Problem found!
  instance isa (m-non-normative-filler-problem)
  explanation: m-expl1
  path: (events old-age object)
  expected-value: m-old-animal
  observed-value: m-swale
... explanation is not believable.
Checking if explanation is useful for the purpose of prediction.
Is the explanation predictive?
... Explanation is adequate for prediction.
Problem found: m-prob5
Checking seriousness of problems.
... m-expl1 is tweakable

Micro Tweaker – Trying to fix "old age"

```
Micro Tweaker
=============
Looking for candidate tweaks in memory for m-prob5.
... found:
    m-replace-action-use-stereotypes-tweak,
    m-replace-actor-use-stereotypes-tweak,
    m-replace-action-use-causals-tweak.
Trying tweak replace-action-search-through-stereotypes on m-expl1.
Looking for effects to be accounted for.
... found m-dead-state
Looking for scripts for m-swale that include m-dead-state
... no scripts found.
... tweak failed.
Trying tweak replace-actor-search-through-stereotypes on m-expl1.
  [tweak unimplemented]
... tweak failed.
Trying tweak replace-action-search-through-causal-rules on m-expl1.
  [tweak unimplemented]
... tweak failed.
```


46

Micro XP Retriever – Seeking XPs Under "death while young"

```
Micro XP Retriever
==================
Searching for XP's.
Looking for retrieval cues in m-prob5.
... Cue: observed m-swale, rather than m-old-animal.
... no XP's found.
```


47

Micro ACCEPTER – Checking Fixx XP

Micro-Accepter
==========
Evaluating m-expl4 for the purpose of prediction.
The underling XP is:
  mop isa (m-xp)
  observed: m-healthy-animal
  expected: m-unhealthy-animal
  events: ((jog . m-jog-event)
    (exert . m-exert-event)
    (defect . m-heart-defect-state)
    (outcome . m-dead-state))
  causals: ((jog => exert) (defect exert => outcome))
  constraints: (((events jog actor) (events exert actor) (events defect owner)
                 (events outcome object)))
Checking if explanation accounts for m-swale-dead-state.
... appears relevant.

Micro Accepter – Checking Fixx XP

Checking believability.
Problem found!
  instance isa (m-non-normative-filler-problem)
  explanation: m-expl4
  path: (events jog actor)
  expected-value: m-human
  observed-value: m-swale
... explanation is not believable.

Micro Accepter – Checking Fixx XP

Checking if explanation is useful for the purpose of prediction.
Is the explanation predictive?
Problem found!
  instance isa (m-non-knowable-antecedent-problem)
  explanation: m-expl4
  fact: m-heart-defect-state
  event-label: defect
... Explanation m-expl4 is not adequate for prediction.
Problems found:
  m-prob6, m-prob7.
Checking seriousness of problems.
... m-expl4 is tweakable

Micro Tweaker – Tweaking Fixx XP

Micro Tweaker
==========
Looking for candidate tweaks in memory for m-prob6.
... found:
   m-replace-action-use-stereotypes-tweak,
   m-replace-actor-use-stereotypes-tweak,
   m-replace-action-use-causals-tweak.
Trying tweak replace-action-search-through-stereotypes on m-expl4.
Looking for effects to be accounted for.
... found m-exert-event
Looking for scripts for m-swale that include m-exert-event
... found m-horserace-event

Micro Tweaker

Creating a new XP m-xp8 with m-horserace-event instead of m-jog-event
Applying m-xp8.
... generated m-expl9.
... tweak generated: (m-expl9).
Trying tweak replace-actor-search-through-stereotypes on m-expl4.
[tweak unimplemented]
... tweak failed.
Trying tweak replace-action-search-through-causal-rules on m-expl4.
[tweak unimplemented]
... tweak failed.
Looking for candidate tweaks in memory for m-prob7.
... no tweaks found.

Micro XP Retriever – Seek XP for Jogging + animal

Micro XP Retriever
==========
Searching for XP's.
Looking for retrieval cues in m-prob6.
... Cue: observed m-swale, rather than m-human.
Looking for retrieval cues in m-prob7.
... no XP's found.

Micro Accepter – Checking Horse Race-> Heart Attack

```
Micro-Accepter
==============
Evaluating m-expl9 for the purpose of prediction.
The underling XP is:
  mop isa (m-xp)
  observed: m-healthy-animal
  expected: m-unhealthy-animal
  events: ((jog . m-horserace-event)
           (exert . m-exert-event)
           (defect . m-heart-defect-state)
           (outcome . m-dead-state))
  causals: ((jog => exert) (defect exert => outcome))
  constraints: (((events jog actor) (events exert actor) (events defect owner)
(events outcome object)))
Checking if explanation accounts for m-swale-dead-state.
... appears relevant.
Checking believability.
... no problems.
```


Micro Accepter – Is this Explanation Predictive?

```
Checking if explanation is useful for the purpose of prediction.
Is the explanation predictive?
Problem found!
  instance isa (m-non-knowable-antecedent-problem)
  explanation: m-expl9
  fact: m-heart-defect-state
  event-label: defect
... Explanation m-expl9 is not adequate for prediction.
Problem found: m-prob10
Checking seriousness of problems.
... m-expl9 is acceptable
>
```

