Massachusetts
Institute of
Technology

CSAIL

Improving Information Extraction by Acquiring External Evidence with Reinforcement Learning

Karthik Narasimhan, Adam Yala, Regina Barzilay

CSAIL, MIT

Information Extraction: State of the Art

- Dependence on large training sets

ACE: 300K words

Freebase: 24M relations

Not available for many domains (ex. medicine, crime)

- Even large corpora do not guarantee high performance
  ~ 75% F1 on relation extraction (ACE)
  ~ 58% F1 on event extraction (ACE)

A hard reading task for you

Task: Identify food carcinogens

Coffee significantly reduced ER and cyclin D1 abundance in ER(+) cells …
Coffee reduced the pAkt levels in both ER(+) and ER(-) cells.

A hard reading task for you

Task: Identify food carcinogens

> **Coffee** significantly reduced ER and cyclin D1 abundance in ER(+) cells …
> **Coffee** reduced the pAkt levels in both ER(+) and ER(-) cells.

Is coffee a carcinogen?

A hard reading task for machines: IE

Extraction (NumWounded)

CBS Chicago

A 2 year old girl and four other people were wounded in a shooting in West Englewood Thursday night, police said

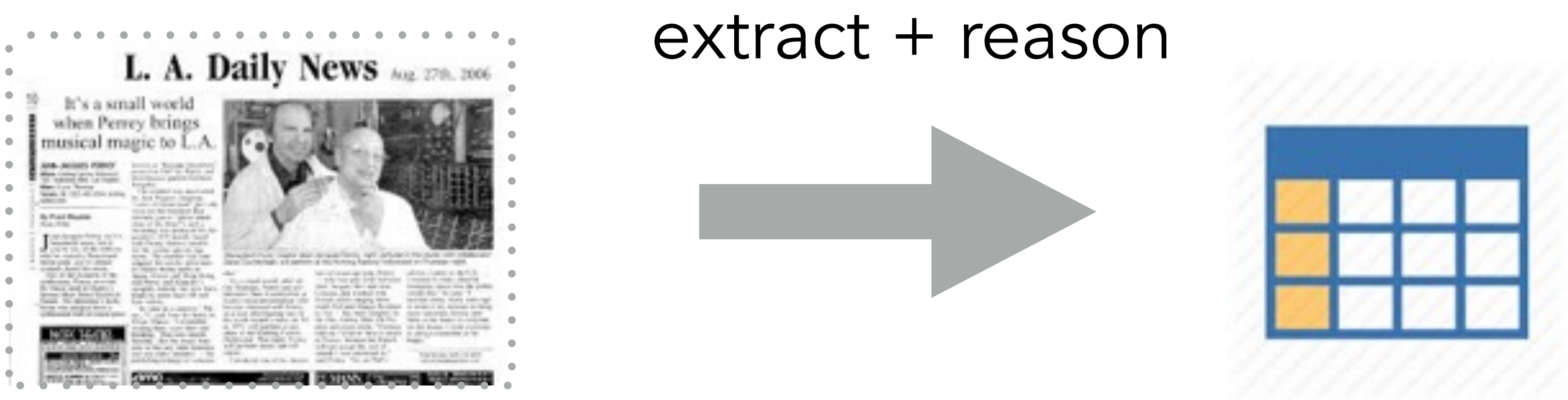
four ✗

A hard reading task: IE (not always!)

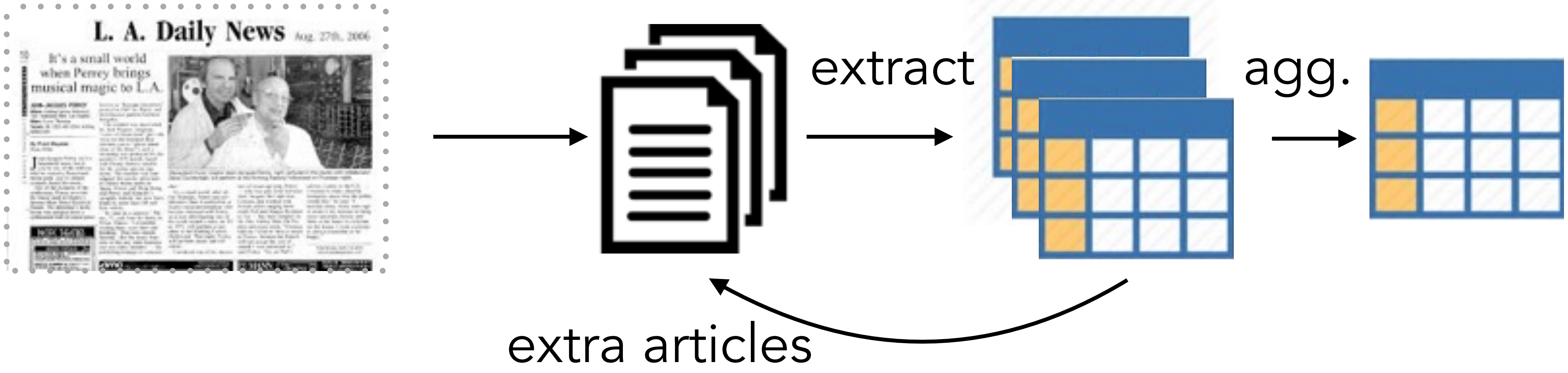
		Extraction (NumWounded)
CBS Chicago	A 2 year old girl and four other people were wounded in a shooting in West Englewood Thursday night, police said	four ✗
abc 7	The last shooting left five people wounded.	five ✓

Incorporate External Evidence

Traditional formulation

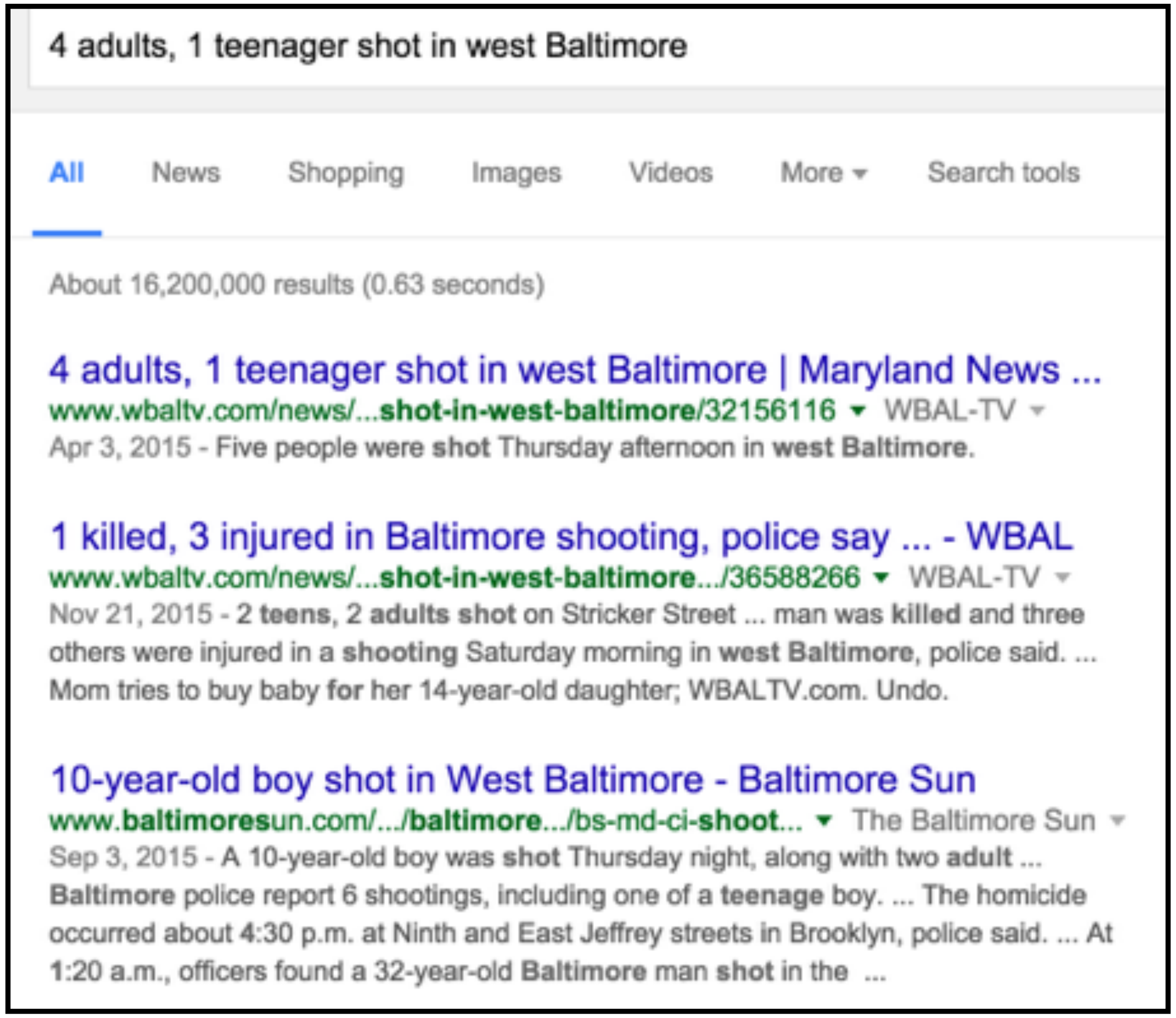

L. A. Daily News Aug. 27th, 2006
It's a small world when Perrey brings musical magic to L.A.
extract + reason

Our approach


L. A. Daily News Aug. 27th, 2006
It's a small world when Perrey brings musical magic to L.A.
extract
agg.
extra articles

Challenges

1. Event Coreference



4 adults, 1 teenager shot in west Baltimore

About 16,200,000 results (0.63 seconds)

4 adults, 1 teenager shot in west Baltimore | Maryland News ...
www.wbaltv.com/news/...shot-in-west-baltimore/32156116 ▾ WBAL-TV ▾
Apr 3, 2015 - Five people were shot Thursday afternoon in west Baltimore.

1 killed, 3 injured in Baltimore shooting, police say ... - WBAL
www.wbaltv.com/news/...shot-in-west-baltimore.../36588266 ▾ WBAL-TV ▾
Nov 21, 2015 - 2 teens, 2 adults shot on Stricker Street ... man was killed and three others were injured in a shooting Saturday morning in west Baltimore, police said. ... Mom tries to buy baby for her 14-year-old daughter; WBALTV.com. Undo.

10-year-old boy shot in West Baltimore - Baltimore Sun
www.baltimoresun.com/.../baltimore.../bs-md-ci-shoot... ▾ The Baltimore Sun ▾
Sep 3, 2015 - A 10-year-old boy was shot Thursday night, along with two adult ... Baltimore police report 6 shootings, including one of a teenage boy. ... The homicide occurred about 4:30 p.m. at Ninth and East Jeffrey streets in Brooklyn, police said. ... At 1:20 a.m., officers found a 32-year-old Baltimore man shot in the ...

Several irrelevant articles!

2. Reconciling Predictions

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

Shooter: Scott Westerhuis

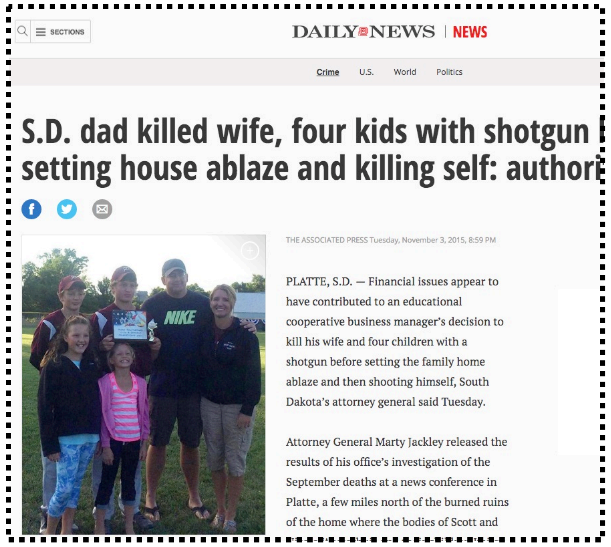
NumKilled: 6

Location: Platte

Inconsistent extractions

Learning through Reinforcement

Original



S.D. dad killed wife, four kids with shotgun setting house ablaze and killing self: authori

THE ASSOCIATED PRESS Tuesday, November 3, 2015, 8:59 PM

PLATTE, S.D. — Financial issues appear to have contributed to an educational cooperative business manager's decision to kill his wife and four children with a shotgun before setting the family home ablaze and then shooting himself, South Dakota's attorney general said Tuesday.

Attorney General Marty Jackley released the results of his office's investigation of the September deaths at a news conference in Platte, a few miles north of the burned ruins of the home where the bodies of Scott and

extract →

Shooter: Scott Westerhuis

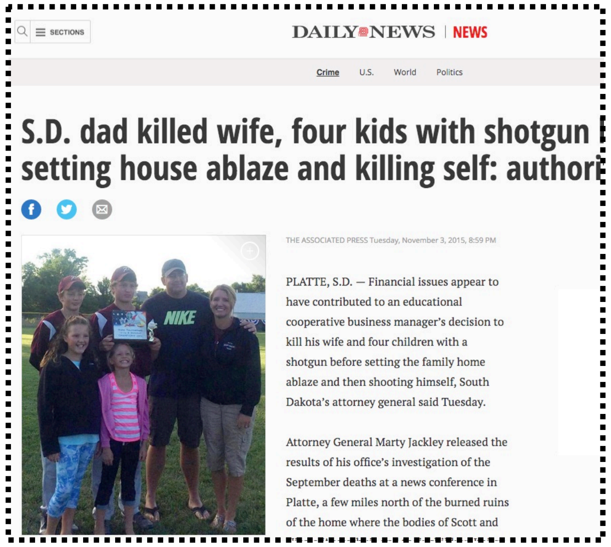
NumKilled: 4

Location: S.D

Start with traditional extraction system

Learning through Reinforcement

Original



SECTIONS

DAILY NEWS | NEWS

Crime U.S. World Politics

S.D. dad killed wife, four kids with shotgun
setting house ablaze and killing self: authori

THE ASSOCIATED PRESS Tuesday, November 3, 2015, 8:59 PM

PLATTE, S.D. — Financial issues appear to have contributed to an educational cooperative business manager's decision to kill his wife and four children with a shotgun before setting the family home ablaze and then shooting himself, South Dakota's attorney general said Tuesday.

Attorney General Marty Jackley released the results of his office's investigation of the September deaths at a news conference in Platte, a few miles north of the burned ruins of the home where the bodies of Scott and

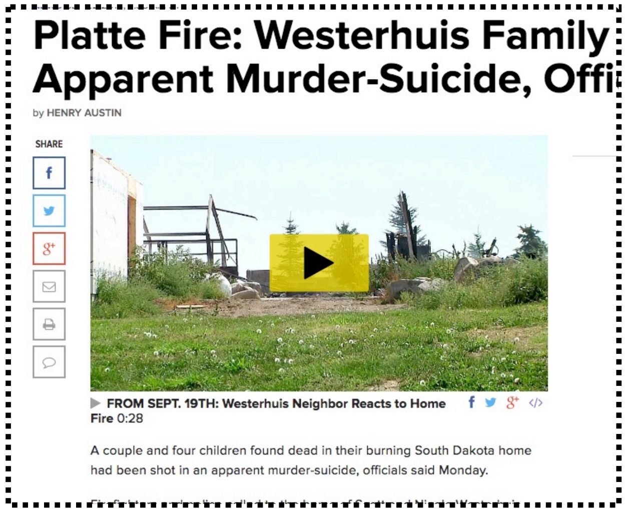
extract

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

query



Platte Fire: Westerhuis Family
Apparent Murder-Suicide, Offi

by HENRY AUSTIN

SHARE

FROM SEPT. 19TH: Westerhuis Neighbor Reacts to Home Fire 0:28

A couple and four children found dead in their burning South Dakota home had been shot in an apparent murder-suicide, officials said Monday.

extract

Shooter: Scott Westerhuis

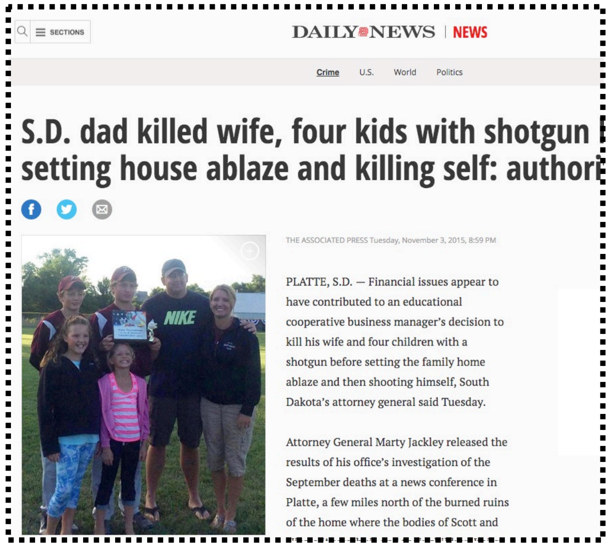
NumKilled: 6

Location: Platte

Perform a query and extract from a new article

Learning through Reinforcement

Original



SECTIONS

DAILY NEWS | NEWS

Crime U.S. World Politics

S.D. dad killed wife, four kids with shotgun
setting house ablaze and killing self: authori

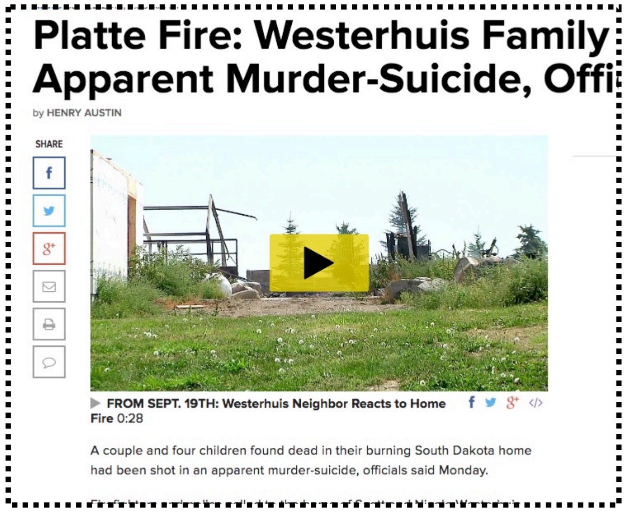
THE ASSOCIATED PRESS Tuesday, November 3, 2015, 8:59 PM

PLATTE, S.D. — Financial issues appear to have contributed to an educational cooperative business manager's decision to kill his wife and four children with a shotgun before setting the family home ablaze and then shooting himself, South Dakota's attorney general said Tuesday.

Attorney General Marty Jackley released the results of his office's investigation of the September deaths at a news conference in Platte, a few miles north of the burned ruins of the home where the bodies of Scott and

extract

search



Platte Fire: Westerhuis Family
Apparent Murder-Suicide, Offi

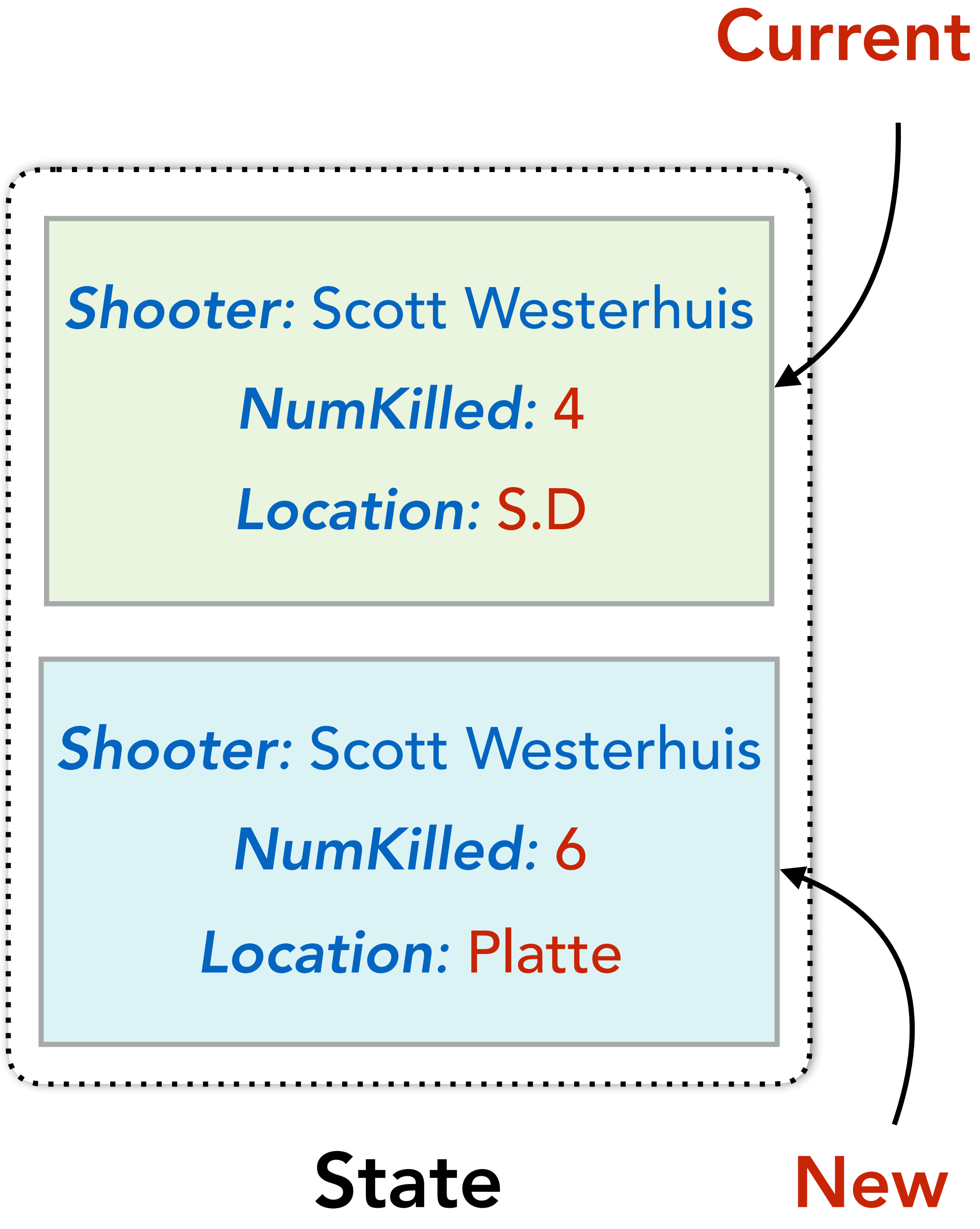
by HENRY AUSTIN

SHARE

FROM SEPT. 19TH: Westerhuis Neighbor Reacts to Home Fire 0:28

A couple and four children found dead in their burning South Dakota home had been shot in an apparent murder-suicide, officials said Monday.

extract



Current

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

Shooter: Scott Westerhuis

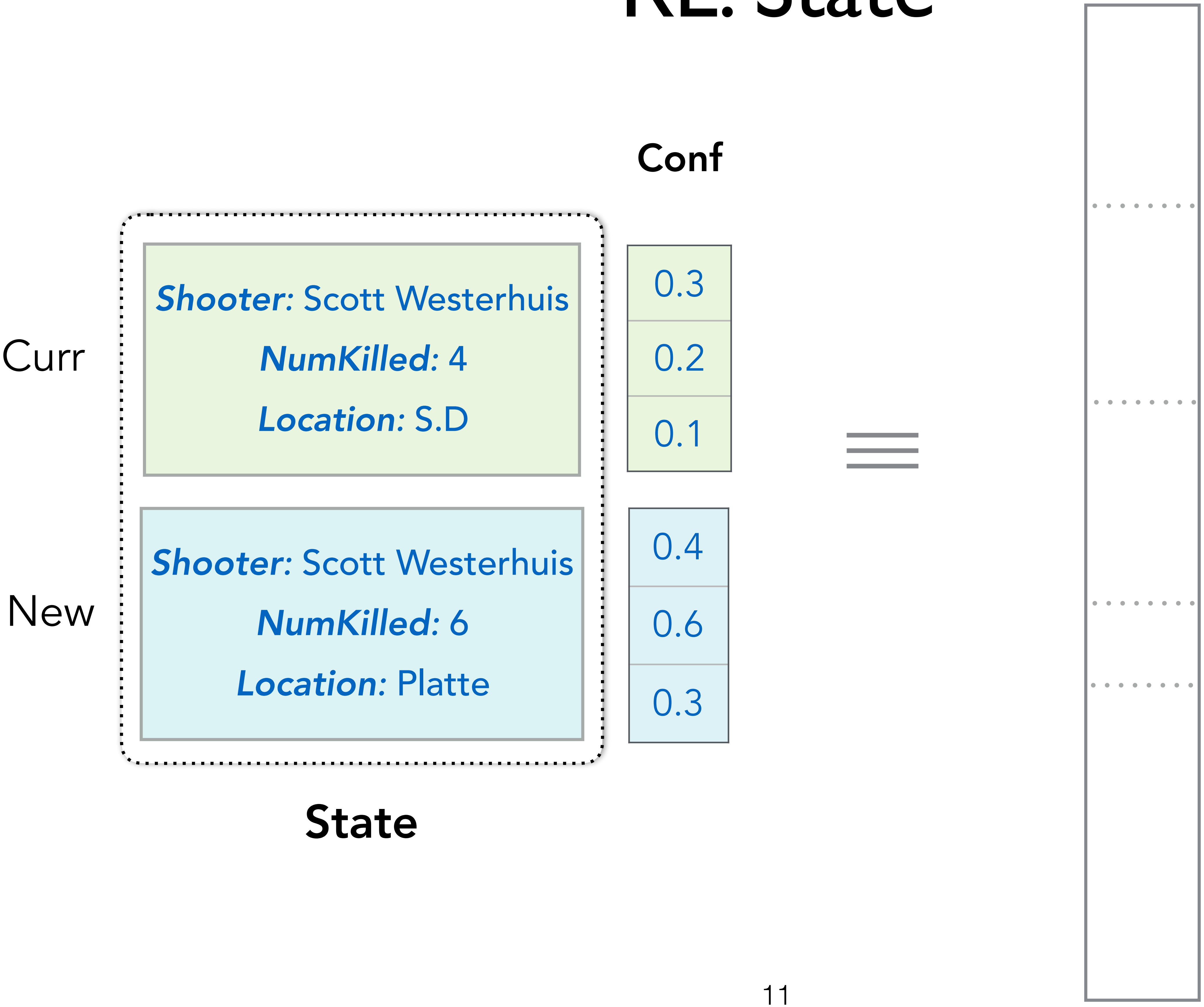
NumKilled: 6

Location: Platte

State

New

RL: State



Conf

Curr

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

0.3

0.2

0.1

New

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

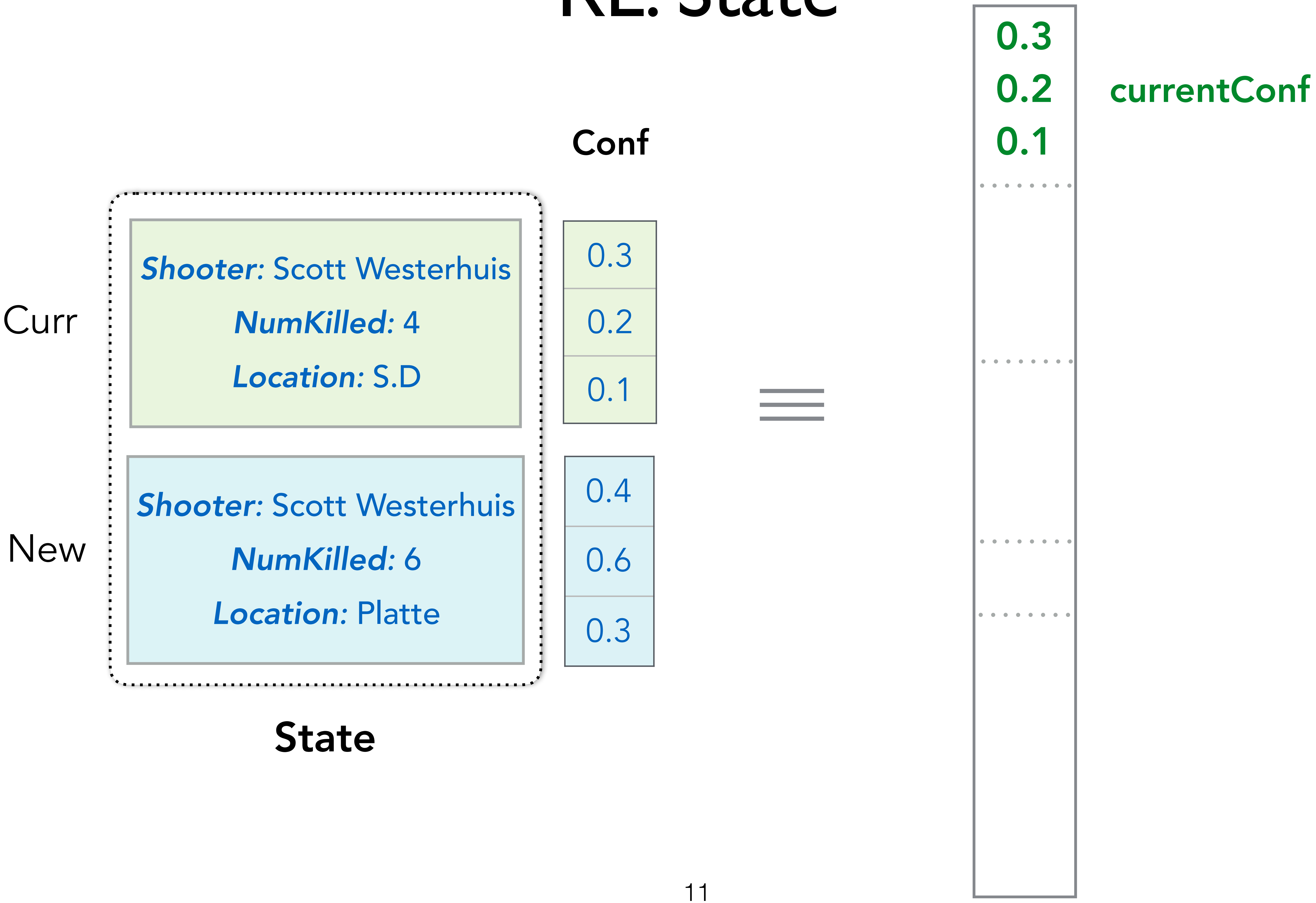
0.4

0.6

0.3

State

RL: State



Conf

Curr

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

0.3
0.2
0.1

New

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

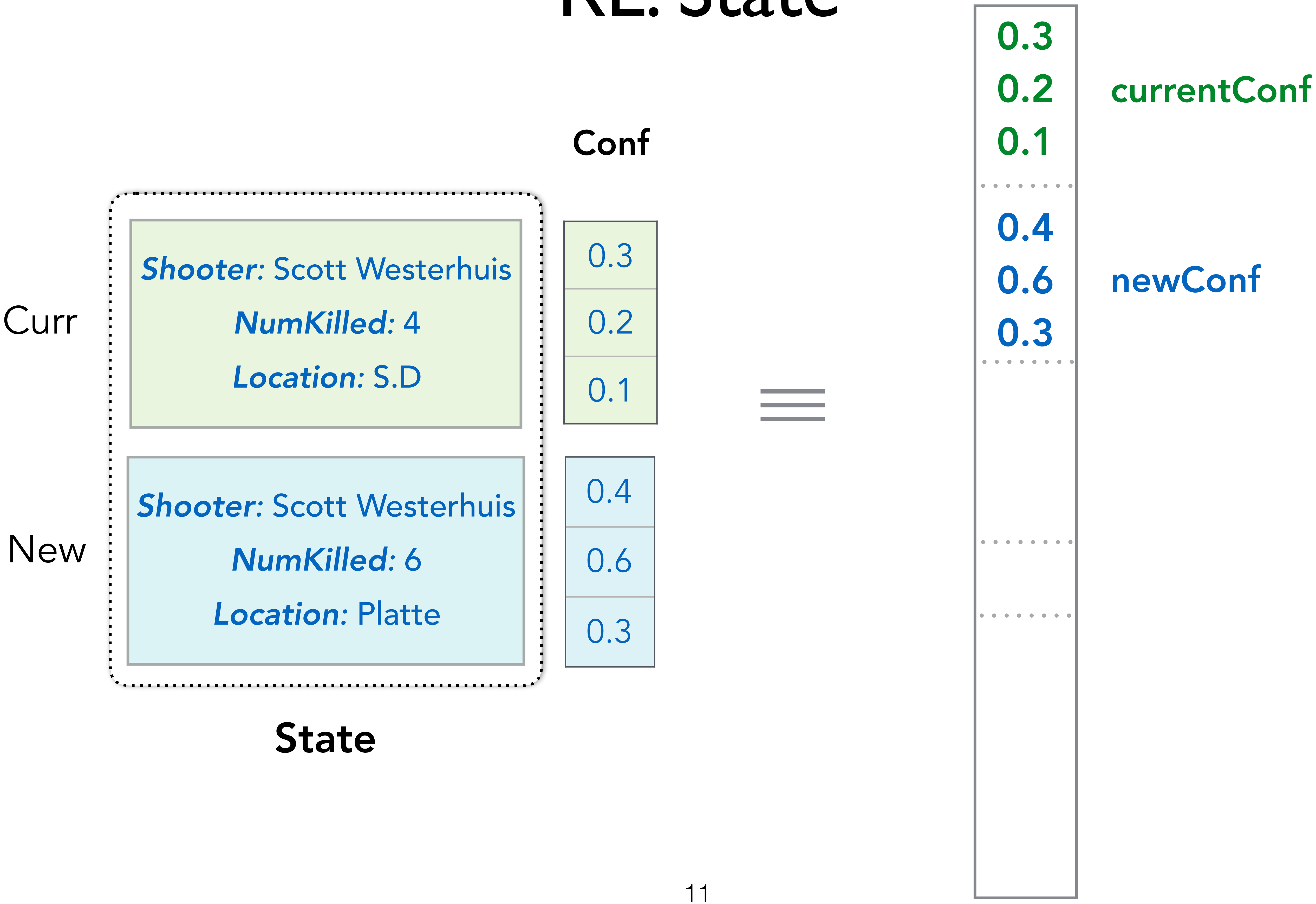
0.4
0.6
0.3

State

0.3
0.2
0.1

currentConf

RL: State



Conf

Curr

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

0.3

0.2

0.1

New

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

0.4

0.6

0.3

State

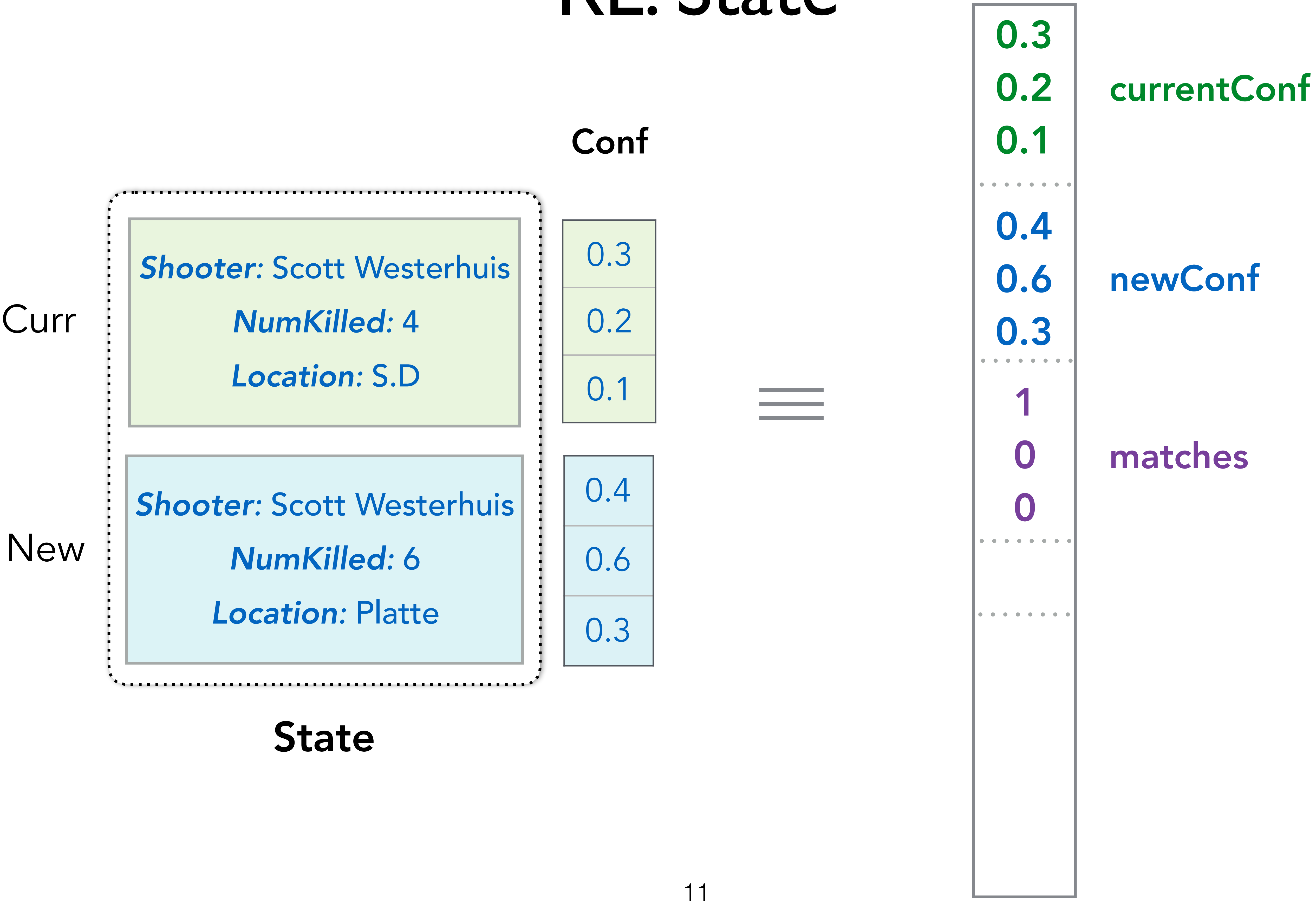
0.3
0.2
0.1

currentConf

0.4
0.6
0.3

newConf

RL: State



Conf

Curr

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

0.3

0.2

0.1

New

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

0.4

0.6

0.3

State

0.3

0.2

0.1

currentConf

0.4

0.6

0.3

newConf

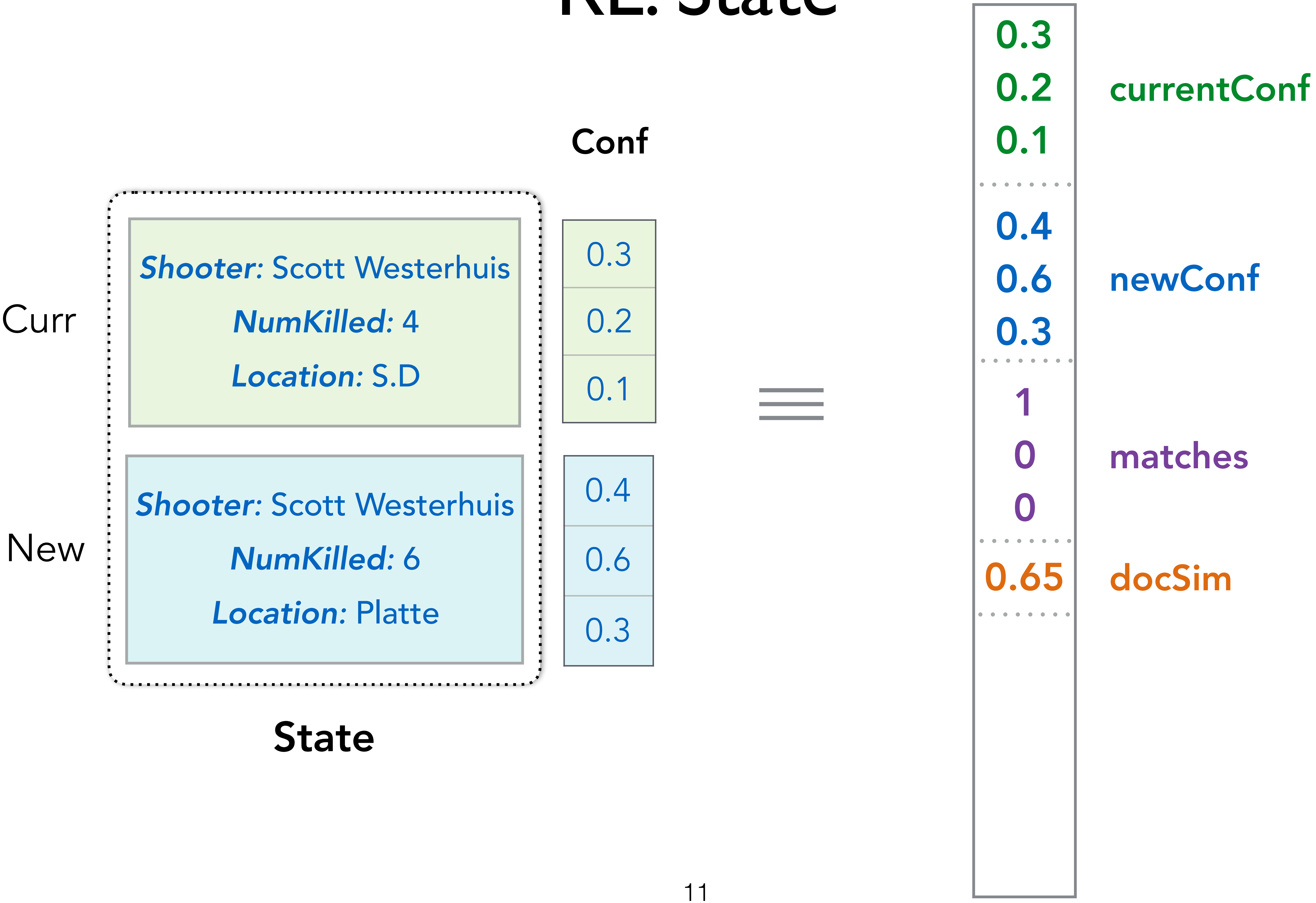
1

0

0

matches

RL: State



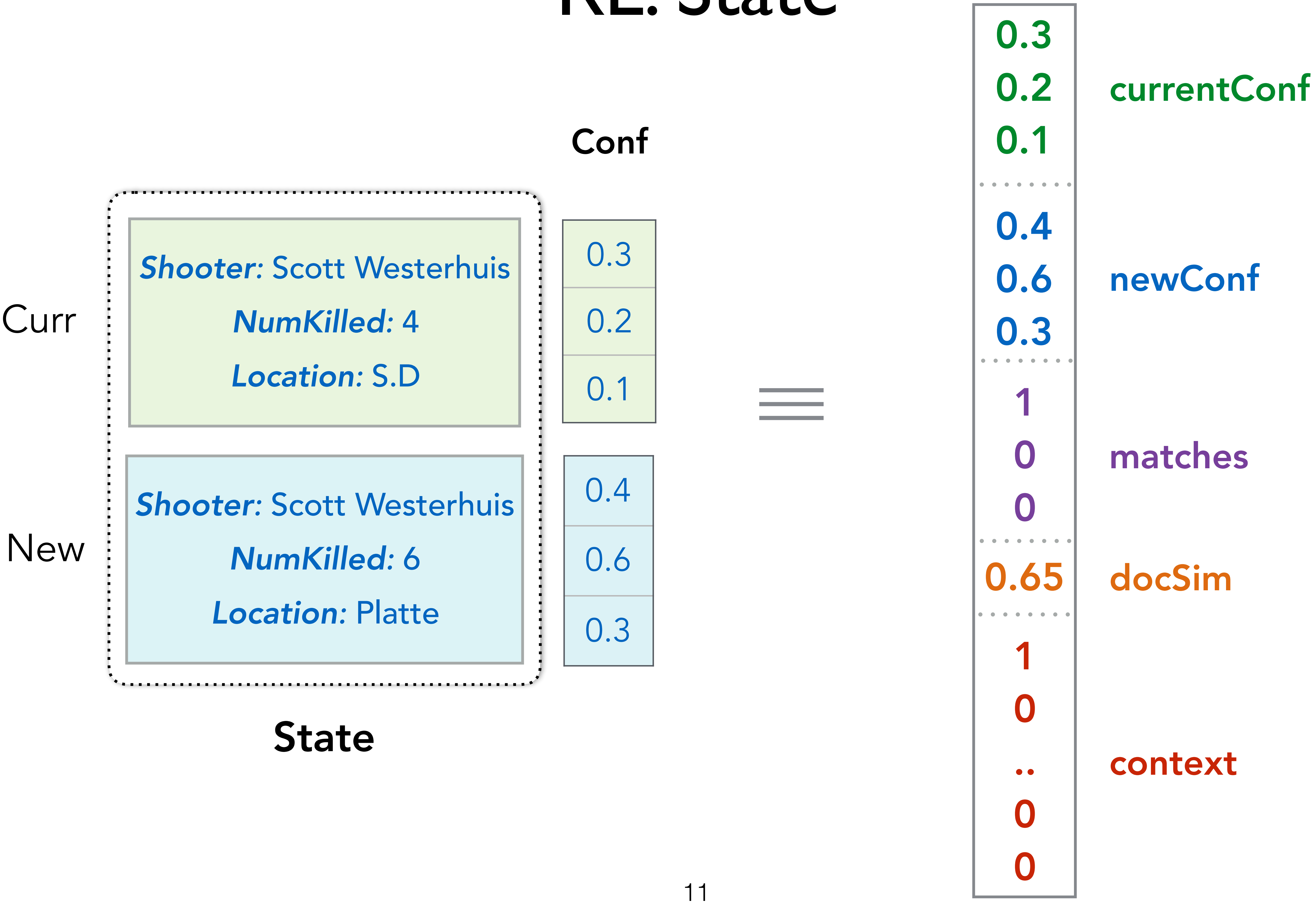
		Conf
Curr	***Shooter:*** Scott Westerhuis	0.3
	NumKilled: 4	0.2
	Location: S.D	0.1
New	***Shooter:*** Scott Westerhuis	0.4
	NumKilled: 6	0.6
	Location: Platte	0.3

State

≡

0.3	
0.2	currentConf
0.1	
0.4	
0.6	newConf
0.3	
1	
0	matches
0	
0.65	docSim

RL: State



Conf

Curr

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

Conf
0.3
0.2
0.1

New

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

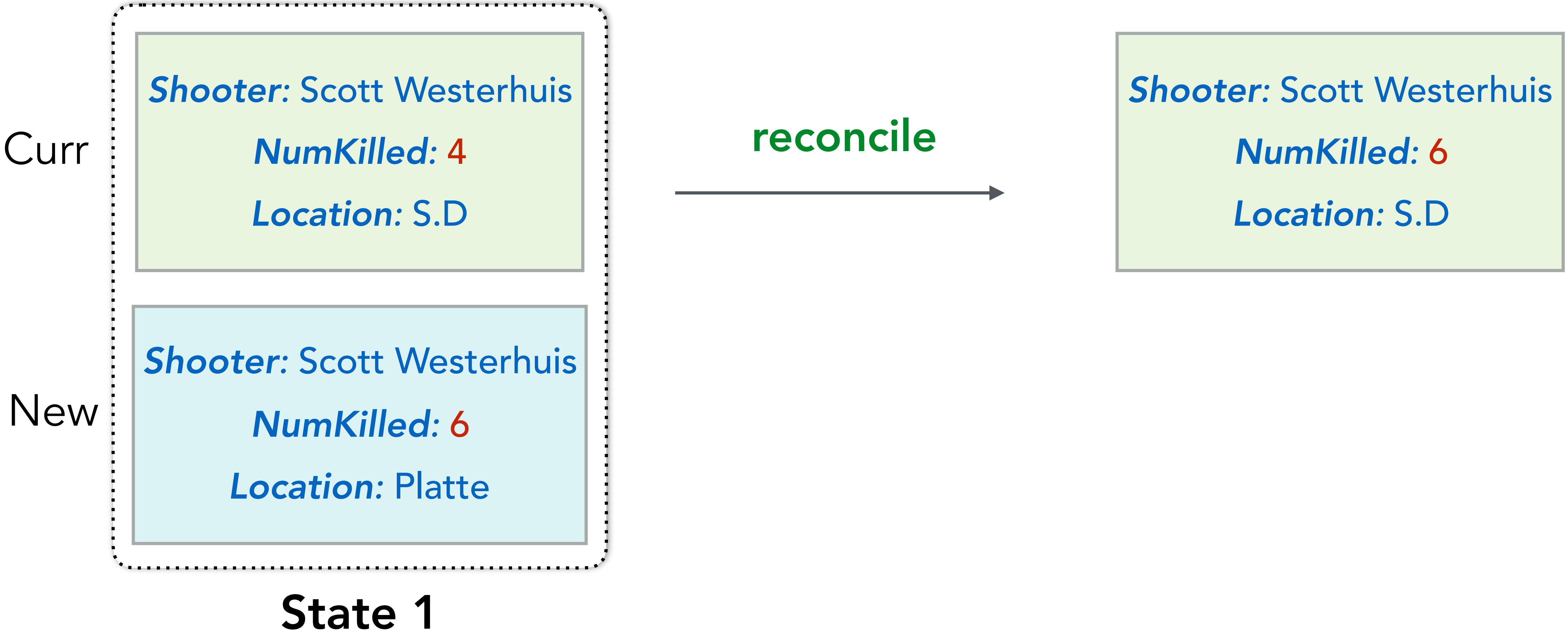
Conf
0.4
0.6
0.3

State

≡

0.3 0.2 0.1	currentConf
0.4 0.6 0.3	newConf
1 0 0	matches
0.65	docSim
1 0 .. 0 0	context

RL: Actions



Curr

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

New

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

State 1

reconcile

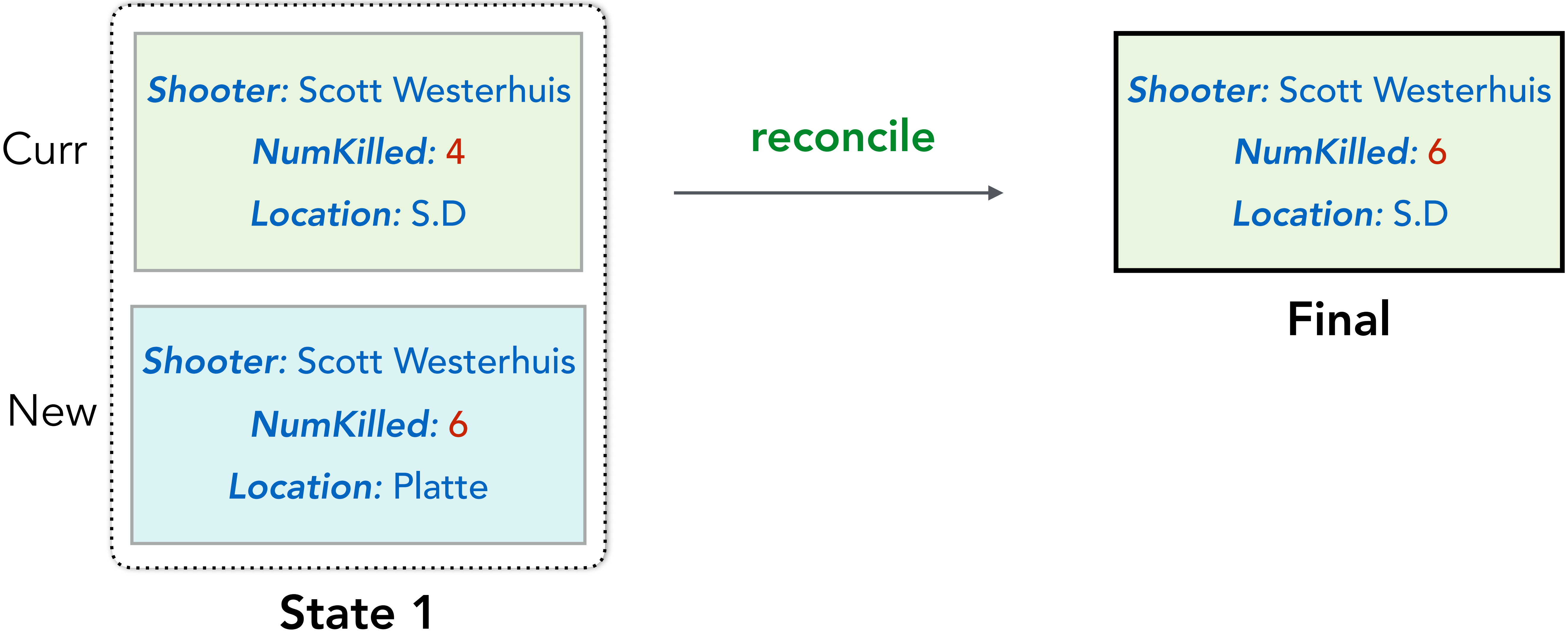
Shooter: Scott Westerhuis

NumKilled: 6

Location: S.D

1. Reconcile (d) old values and new values.
   - Pick a single value, all values or no value from new set

RL: Actions



Curr

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

New

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

State 1

reconcile

Shooter: Scott Westerhuis

NumKilled: 6

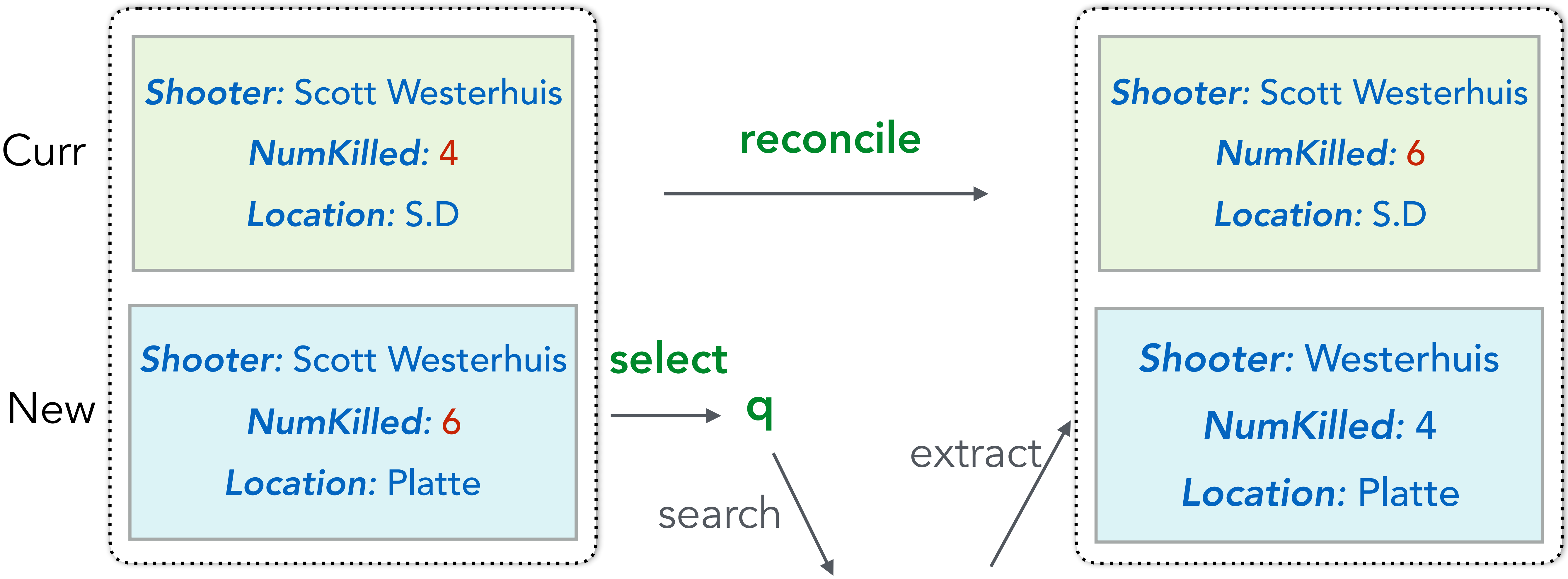
Location: S.D

Final

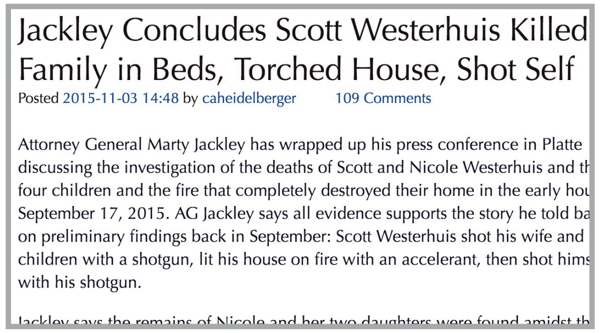
2. Decide how to proceed:

- **Stop**

RL: Actions



State 1


Jackley Concludes Scott Westerhuis Killed Family in Beds, Torched House, Shot Self

Posted 2015-11-03 14:48 by caheidelberger 109 Comments

Attorney General Marty Jackley has wrapped up his press conference in Platte discussing the investigation of the deaths of Scott and Nicole Westerhuis and th four children and the fire that completely destroyed their home in the early hou September 17, 2015. AG Jackley says all evidence supports the story he told ba on preliminary findings back in September: Scott Westerhuis shot his wife and children with a shotgun, lit his house on fire with an accelerant, then shot hims with his shotgun.

Jackley says the remains of Nicole and her two daughters were found amidst th

State 2

2. Decide how to proceed:

- **Select next query (q)**

Queries

Query templates are induced **automatically**

- Title of original article
- Content words having high mutual information with gold values

```
<title>
<title> +  ( suspect | shooter | said | men | arrested | …)
<title> +  ( injured | wounded | victims | shot | … )
```

Rewards

- Change in accuracy

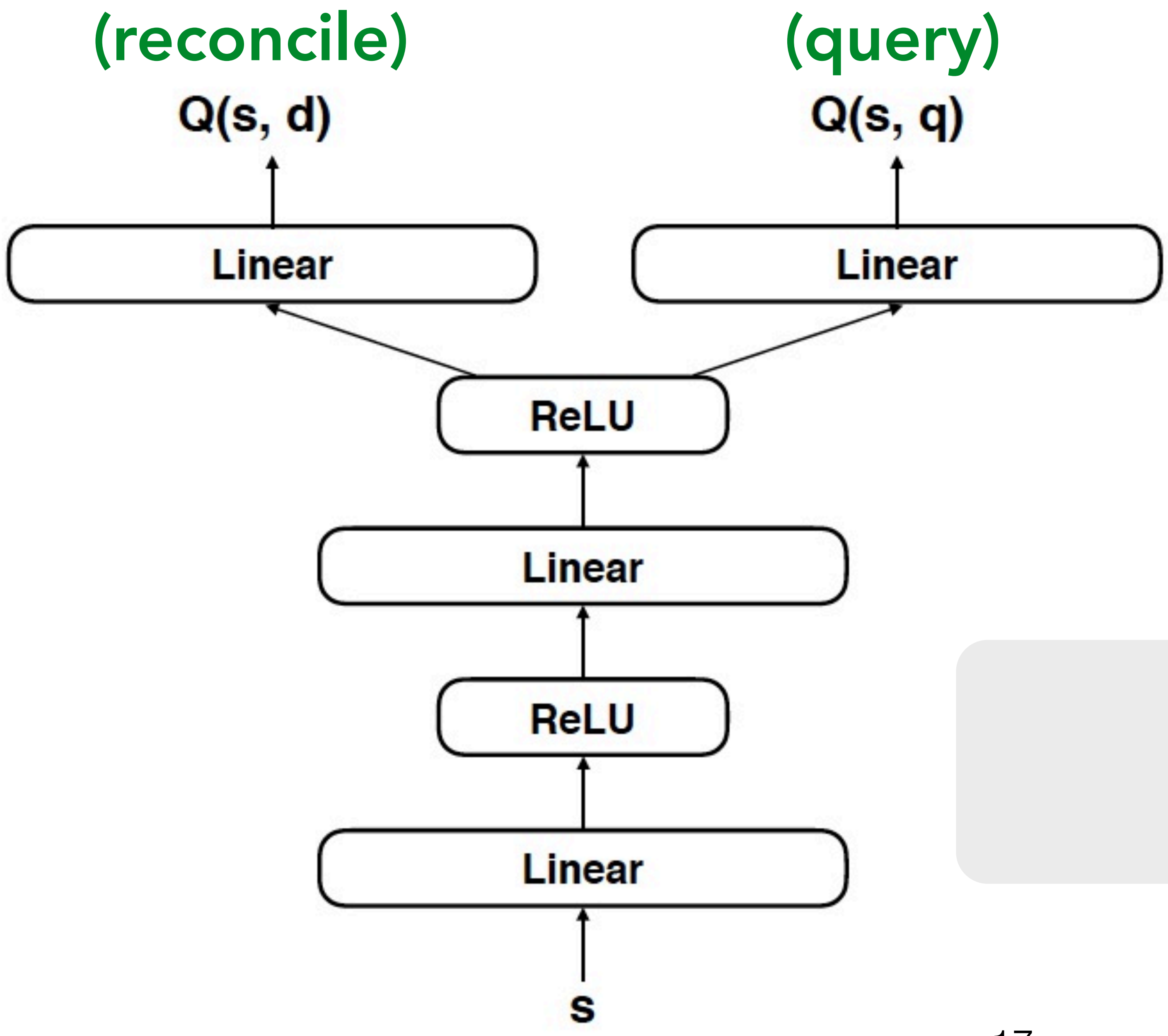
	Previous Values		Current Values
✓	***Shooter:*** Scott Westerhuis	✓	***Shooter:*** Scott Westerhuis
✓	***NumKilled:*** 6	✓	***NumKilled:*** 6
✗	***NumWounded:*** 1	✓	***NumWounded:*** 0
✓	***Location:*** Platte	✓	***Location:*** Platte

$$R(s,a) = \sum_{\text{entity}\, j} \text{Acc}(e^j_{cur}) - \text{Acc}(e^j_{prev}) \quad = 1$$

- Small penalty for each transition

Deep Q-Network

State space is continuous: requires function approximation

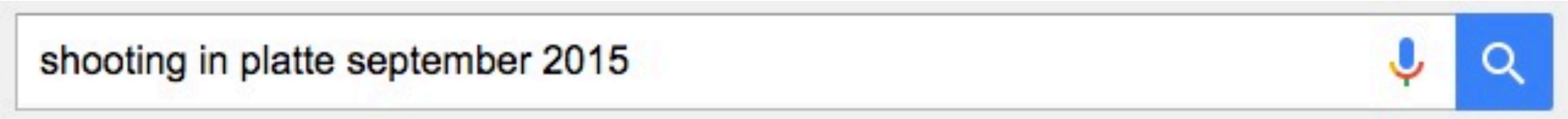


$$Q(s, a) \approx Q(s, a; \theta)$$

Trained to maximize cumulative reward

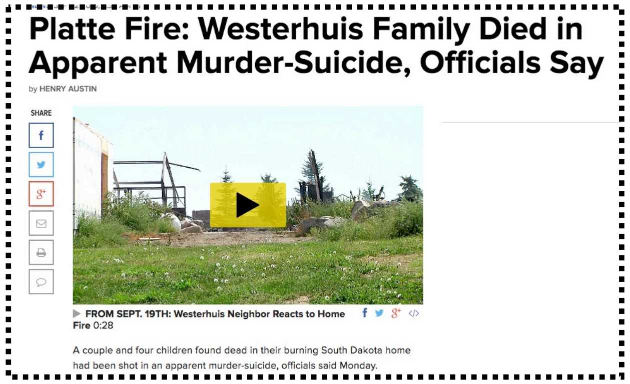
Acquiring External Evidence

1. Select a query to search for articles on the same event



shooting in platte september 2015

2. Use base extractor to obtain values for entities of interest



Platte Fire: Westerhuis Family Died in Apparent Murder-Suicide, Officials Say

by HENRY AUSTIN

SHARE

▶ FROM SEPT. 19TH: Westerhuis Neighbor Reacts to Home Fire 0:28

A couple and four children found dead in their burning South Dakota home had been shot in an apparent murder-suicide, officials said Monday.

extract →

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

3. Reconcile old and new extractions

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

Related Work

- Open Information Extraction *(Etzioni et al., 2011; Fader et al., 2011; Wu and Weld, 2010)*

Related Work

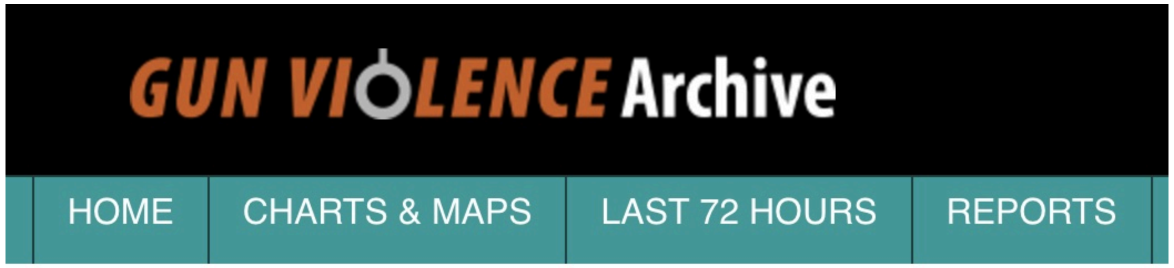
- Open Information Extraction *(Etzioni et al., 2011; Fader et al., 2011; Wu and Weld, 2010)*
- Slot filling *(Surdeanu et al., 2010; Ji and Grishman, 2011)*

Related Work

- Open Information Extraction *(Etzioni et al., 2011; Fader et al., 2011; Wu and Weld, 2010)*
- Slot filling *(Surdeanu et al., 2010; Ji and Grishman, 2011)*
- Searching for additional sources on the web *(Banko et al., 2002, West et al., 2014; Kanani and McCallum, 2012)*

Datasets

1. Mass shootings in the United States



Shooter Name

Num Killed

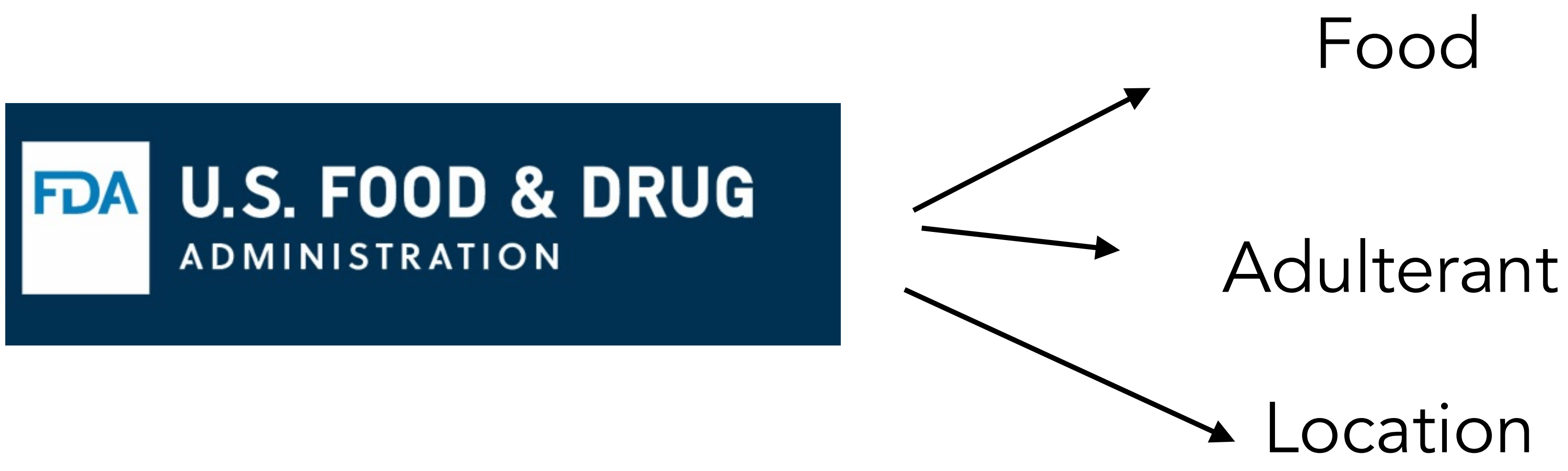
Num Wounded

City

	Train	Test	Dev
Source	306	292	66
Downloaded	8k	7.9k	1.6k

Datasets

2. Adulteration events from Foodshield EMA



Food

Adulterant

Location

	Train	Test	Dev
Source	292	148	42
Downloaded	7.6k	5.3k	1.5k

Base Extraction Model

Maximum entropy model with contextual features

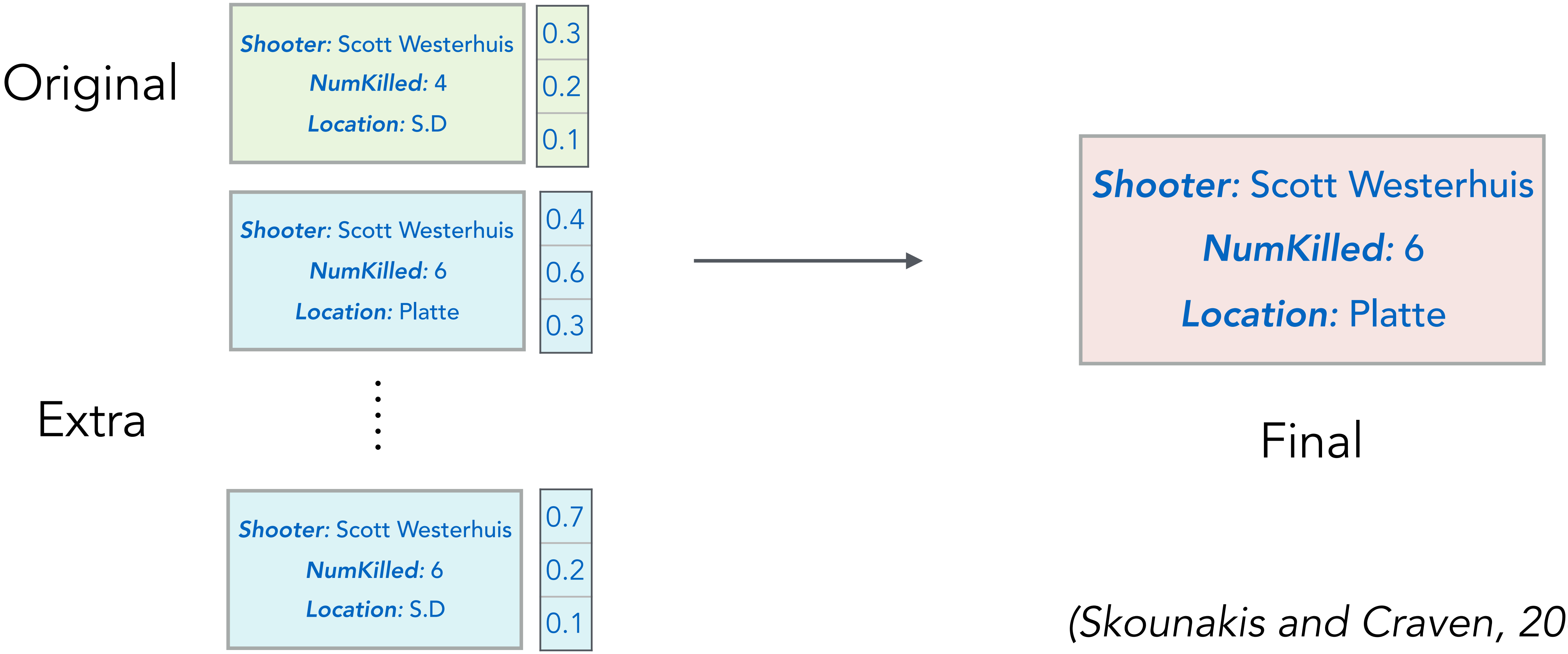
(Chieu and Ng, 2002; Bunescu et al., 2005)

Indirect supervision: Project database values onto articles

Baselines (I)

Simple Aggregation systems:

- *Confidence-based*: Choose entity value with highest confidence



Original

Shooter: Scott Westerhuis
NumKilled: 4
Location: S.D

0.3
0.2
0.1

Shooter: Scott Westerhuis
NumKilled: 6
Location: Platte

0.4
0.6
0.3

Extra

⋮

Shooter: Scott Westerhuis
NumKilled: 6
Location: S.D

0.7
0.2
0.1

→

Shooter: Scott Westerhuis
NumKilled: 6
Location: Platte

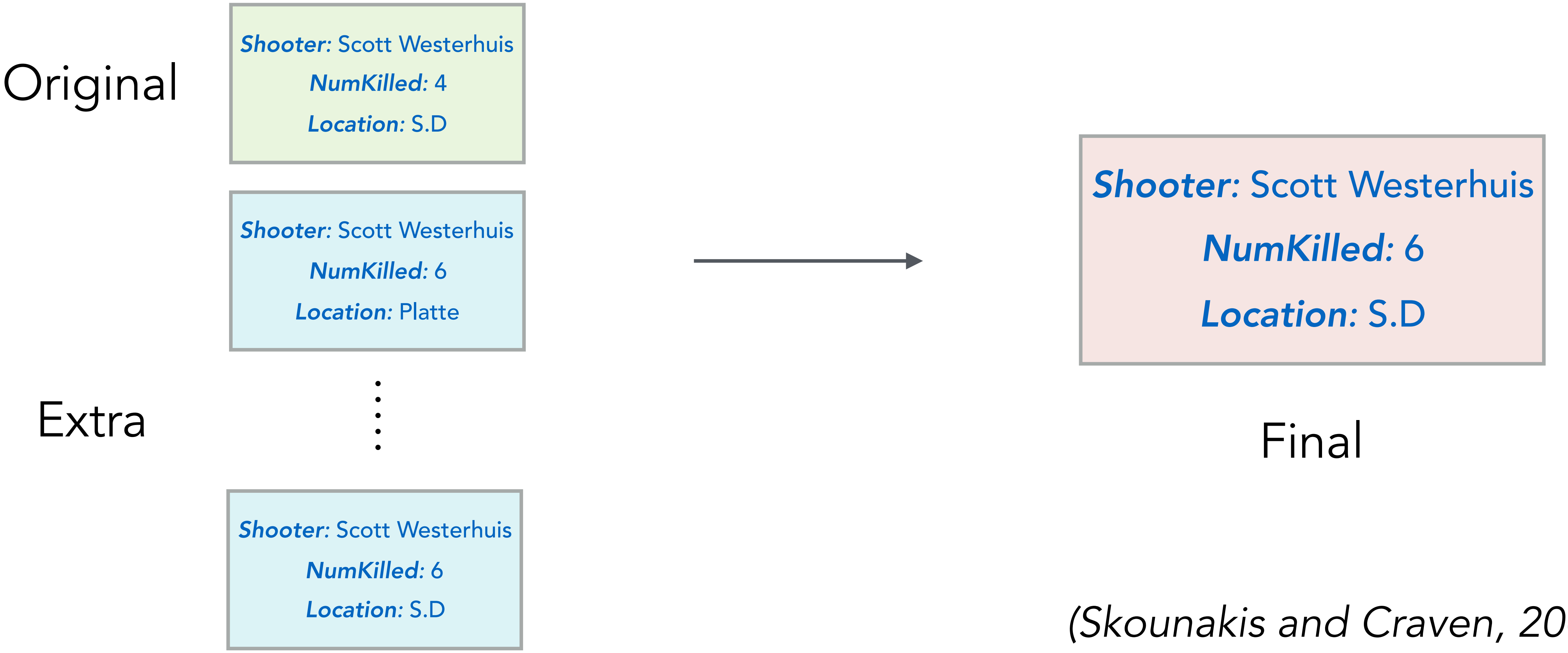
Final

(Skounakis and Craven, 2003)

Baselines (1)

Simple Aggregation systems:

- *Majority-based*: Choose entity value extracted the most from all articles on the event



Original

Shooter: Scott Westerhuis
NumKilled: 4
Location: S.D

Shooter: Scott Westerhuis
NumKilled: 6
Location: Platte

Extra

⋮

Shooter: Scott Westerhuis
NumKilled: 6
Location: S.D

Shooter: Scott Westerhuis
NumKilled: 6
Location: S.D

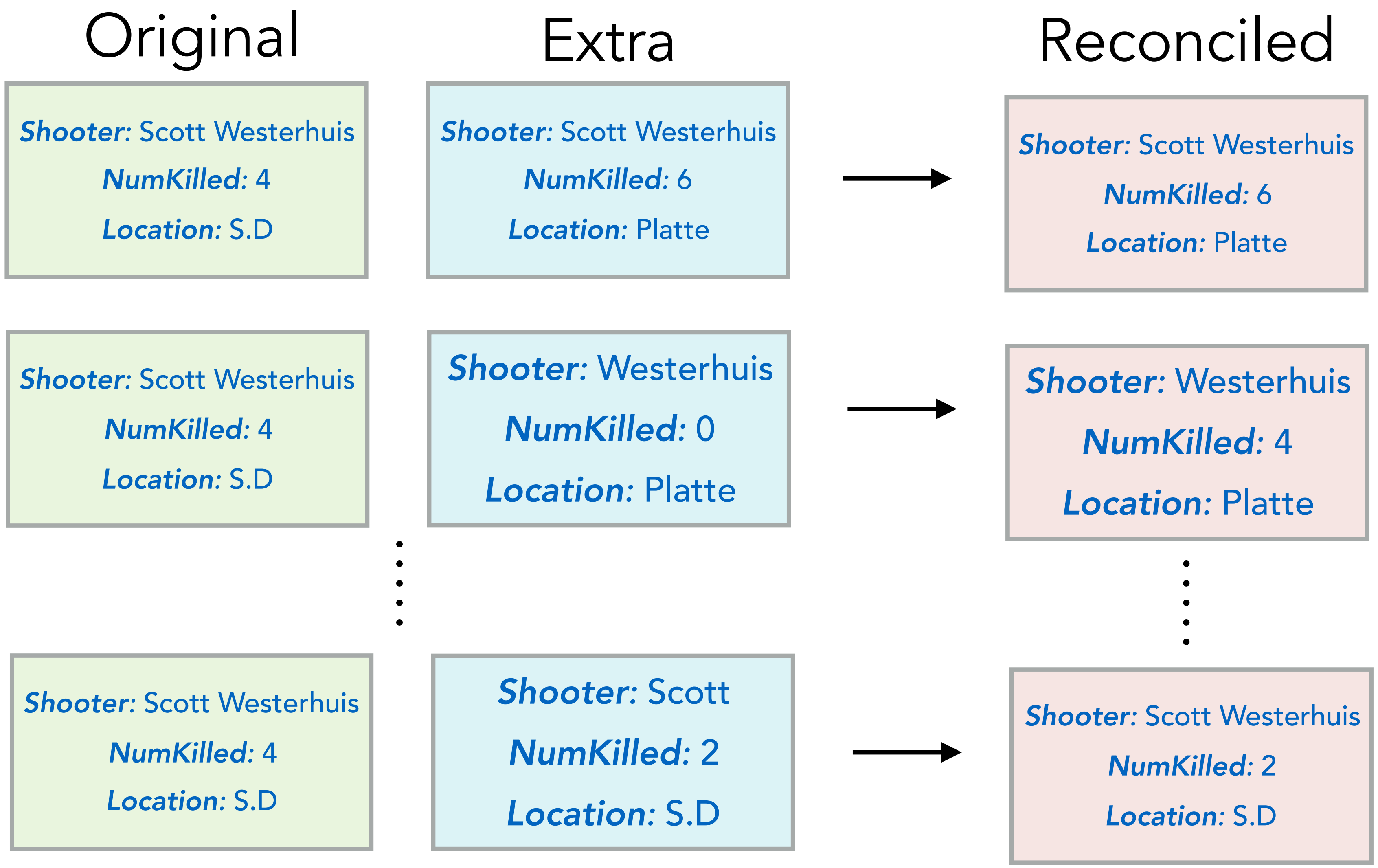
Final

(Skounakis and Craven, 2003)

Baselines (2)

Meta-classifier:

- Same input space S and set of reconciliation decisions as RL agent.

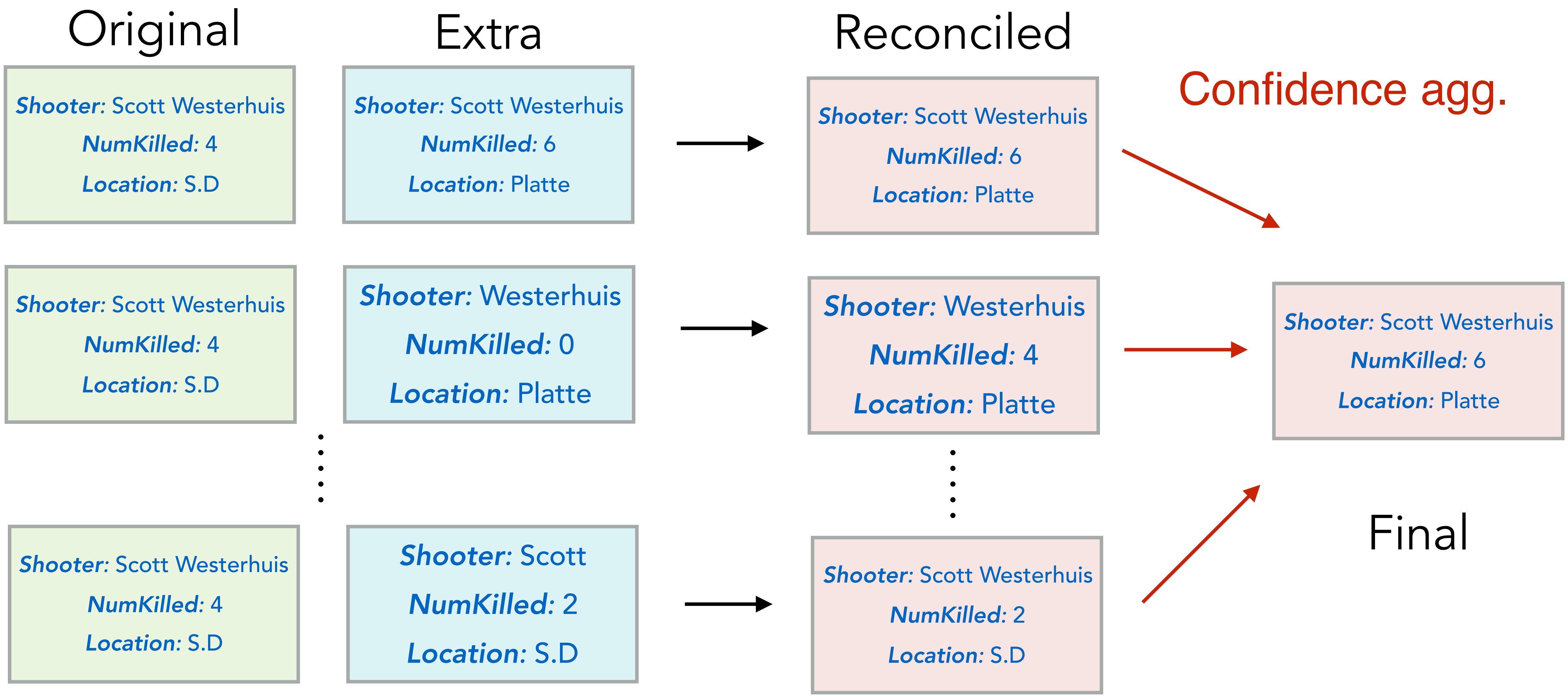


Original	Extra		Reconciled
Shooter: Scott Westerhuis ***NumKilled:*** 4 ***Location:*** S.D	***Shooter:*** Scott Westerhuis ***NumKilled:*** 6 ***Location:*** Platte	→	***Shooter:*** Scott Westerhuis ***NumKilled:*** 6 ***Location:*** Platte
Shooter: Scott Westerhuis ***NumKilled:*** 4 ***Location:*** S.D	***Shooter:*** Westerhuis ***NumKilled:*** 0 ***Location:*** Platte	→	***Shooter:*** Westerhuis ***NumKilled:*** 4 ***Location:*** Platte
⋮			⋮
Shooter: Scott Westerhuis ***NumKilled:*** 4 ***Location:*** S.D	***Shooter:*** Scott ***NumKilled:*** 2 ***Location:*** S.D	→	***Shooter:*** Scott Westerhuis ***NumKilled:*** 2 ***Location:*** S.D

Baselines (2)

Meta-classifier:

- Same input space S and set of reconciliation decisions as RL agent.



Original

Extra

Reconciled

Confidence agg.

Shooter: Scott Westerhuis
NumKilled: 4
Location: S.D

Shooter: Scott Westerhuis
NumKilled: 6
Location: Platte

Shooter: Scott Westerhuis
NumKilled: 6
Location: Platte

Shooter: Scott Westerhuis
NumKilled: 4
Location: S.D

Shooter: Westerhuis
NumKilled: 0
Location: Platte

Shooter: Westerhuis
NumKilled: 4
Location: Platte

Shooter: Scott Westerhuis
NumKilled: 6
Location: Platte

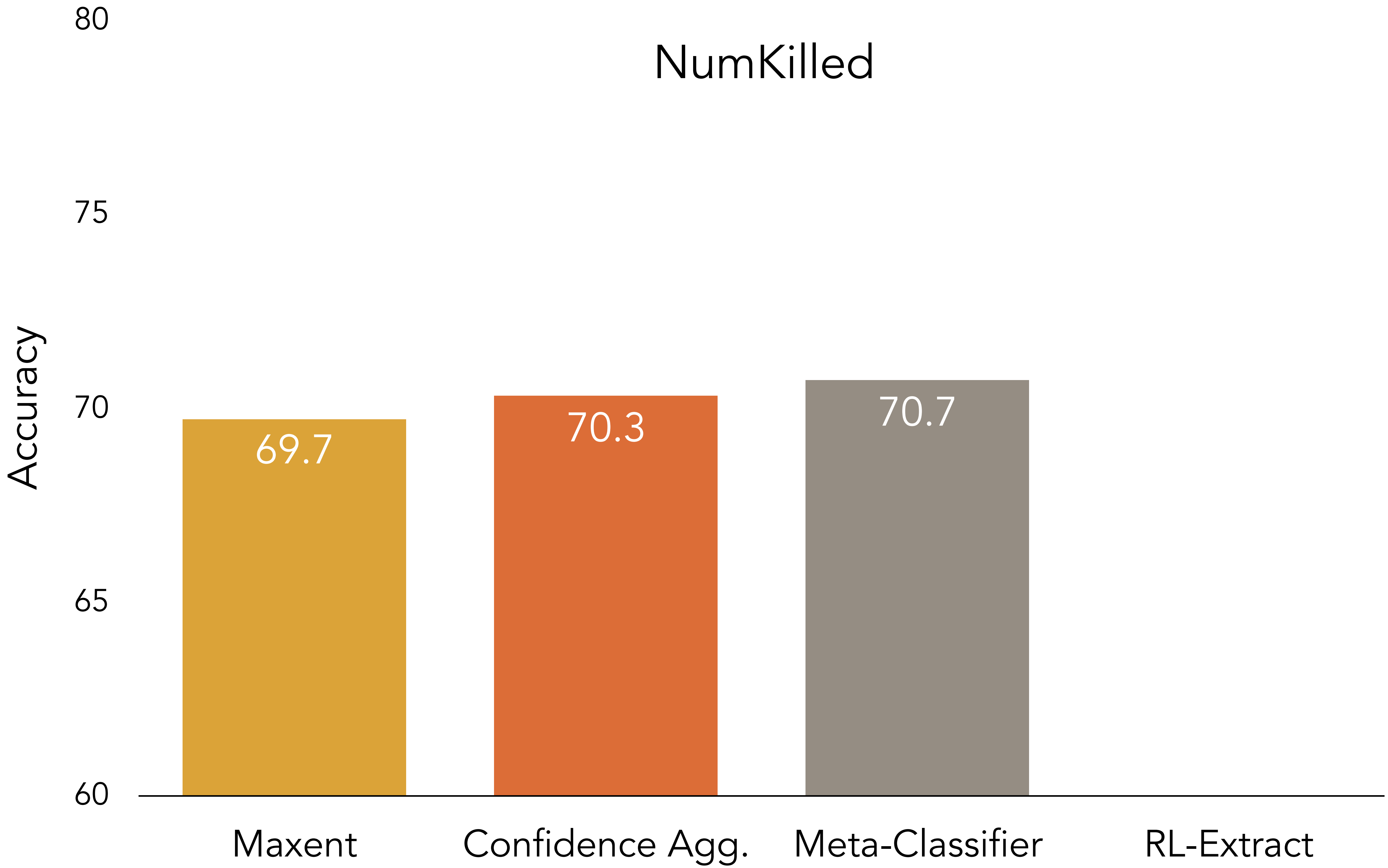
Final

Shooter: Scott Westerhuis
NumKilled: 4
Location: S.D

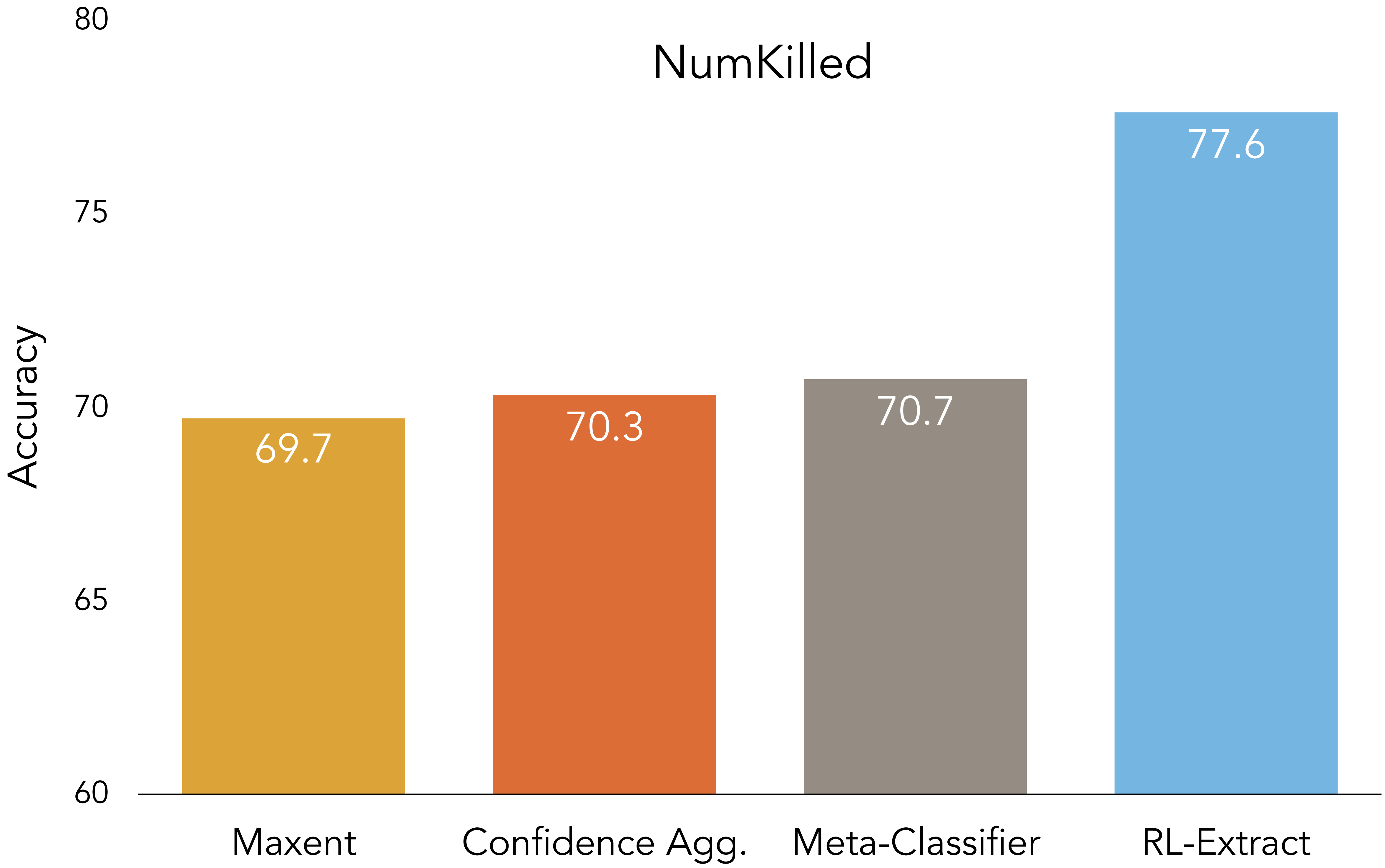
Shooter: Scott
NumKilled: 2
Location: S.D

Shooter: Scott Westerhuis
NumKilled: 2
Location: S.D

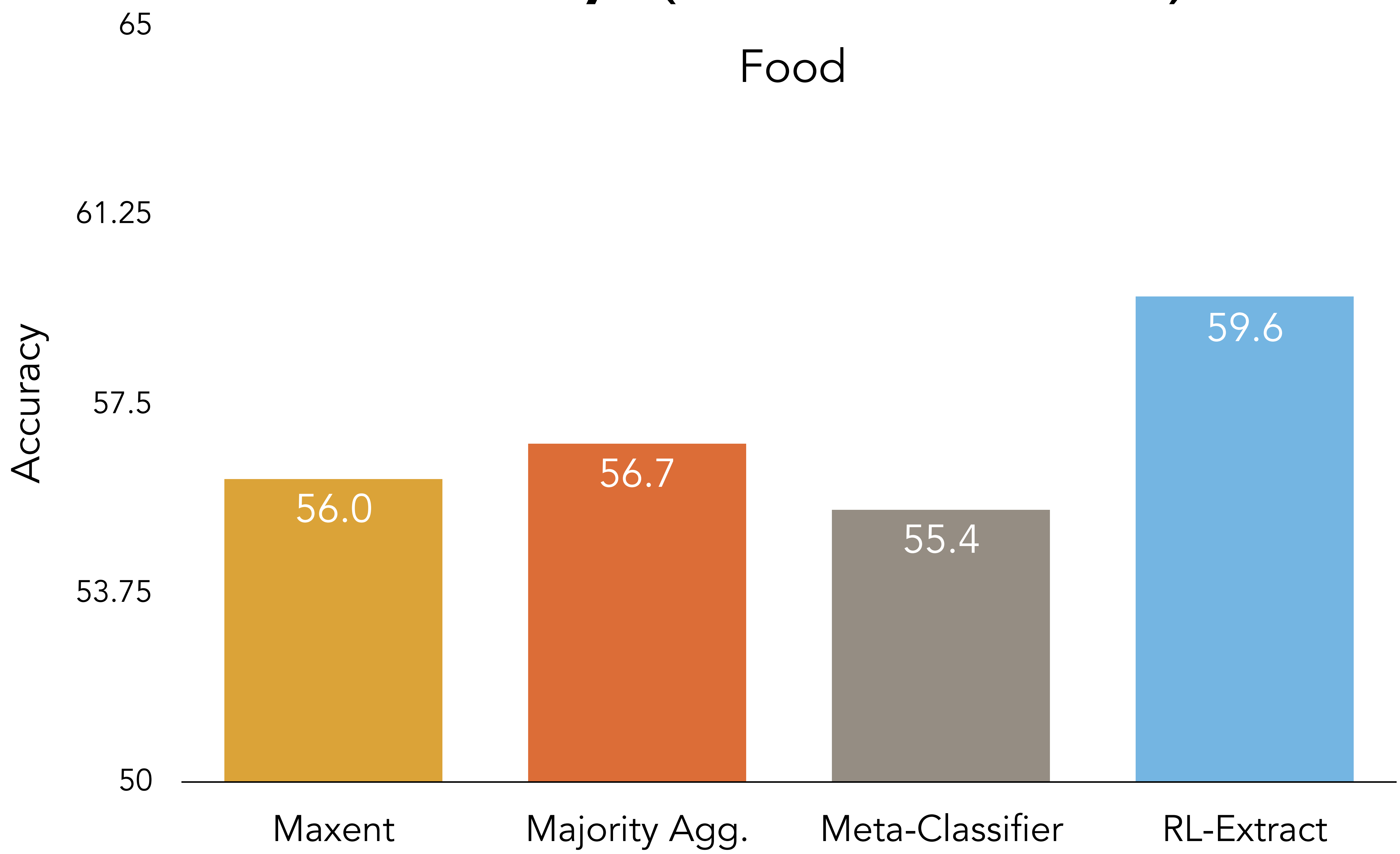
Accuracy (Shootings)



Accuracy (Shootings)



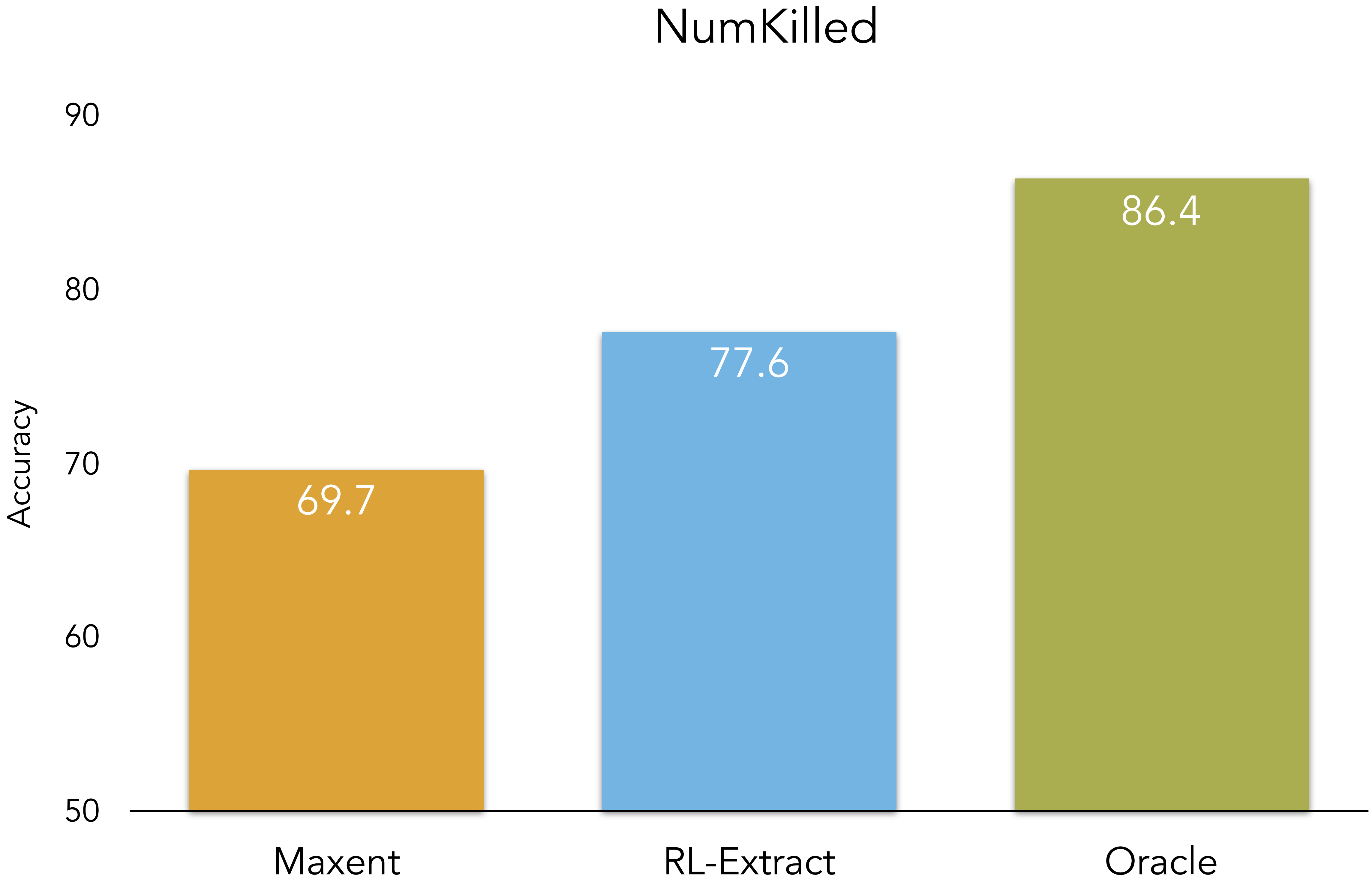
Accuracy (Adulterations)



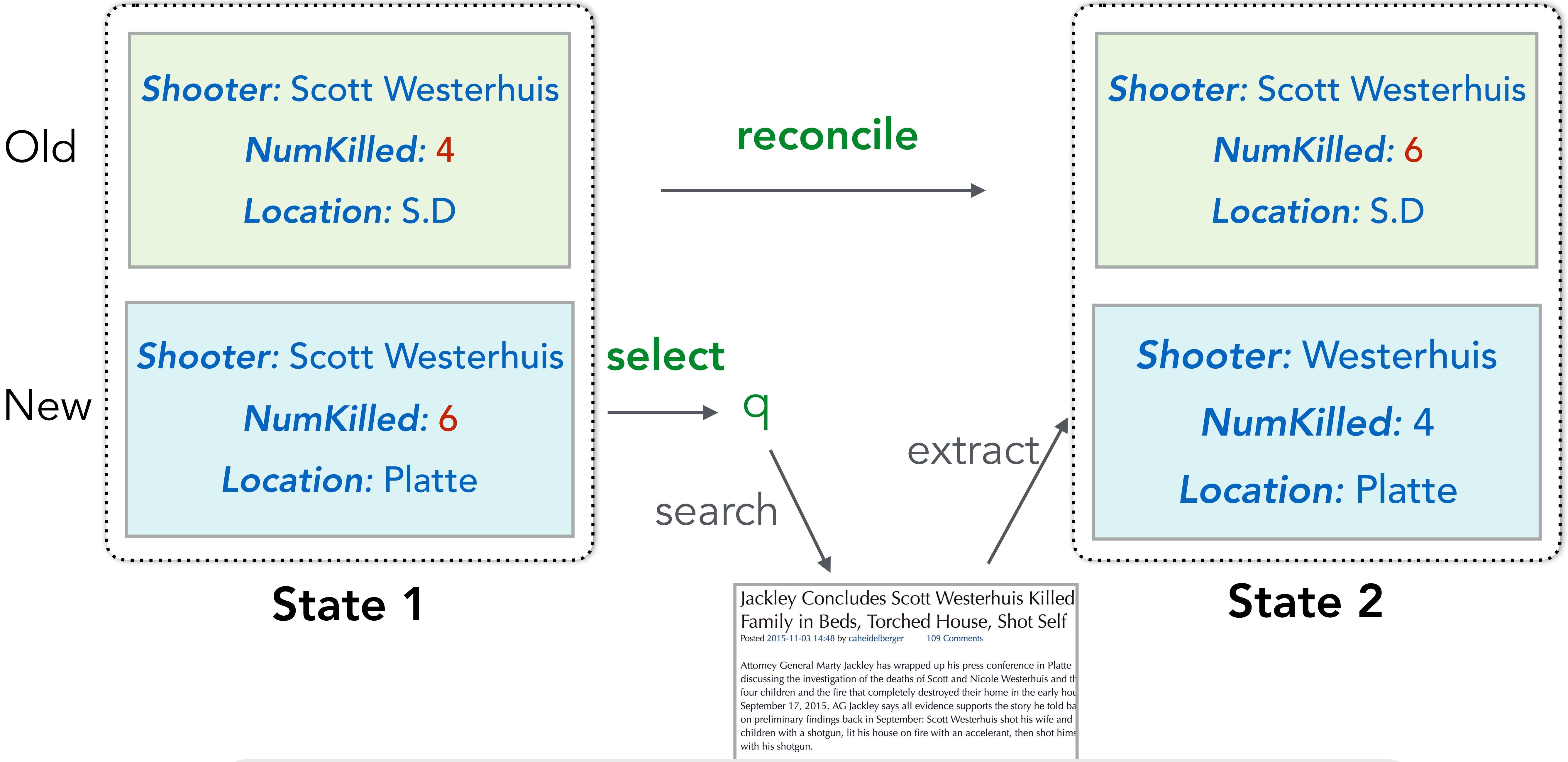
Oracle

- **Given**:
  - Same base extractor
  - Same set of queries
- Agent performing **perfect** reconciliation and querying decisions.
- Upper-bound on performance of any system given these extra articles on each event.

Accuracy (Shootings)



RL-Extract



Old

Shooter: Scott Westerhuis

NumKilled: 4

Location: S.D

New

Shooter: Scott Westerhuis

NumKilled: 6

Location: Platte

State 1

select

q

search

Jackley Concludes Scott Westerhuis Killed Family in Beds, Torched House, Shot Self

Posted 2015-11-03 14:48 by caheidelberger 109 Comments

Attorney General Marty Jackley has wrapped up his press conference in Platte discussing the investigation of the deaths of Scott and Nicole Westerhuis and th four children and the fire that completely destroyed their home in the early hou September 17, 2015. AG Jackley says all evidence supports the story he told ba on preliminary findings back in September: Scott Westerhuis shot his wife and children with a shotgun, lit his house on fire with an accelerant, then shot hims with his shotgun.

extract

reconcile

Shooter: Scott Westerhuis

NumKilled: 6

Location: S.D

Shooter: Westerhuis

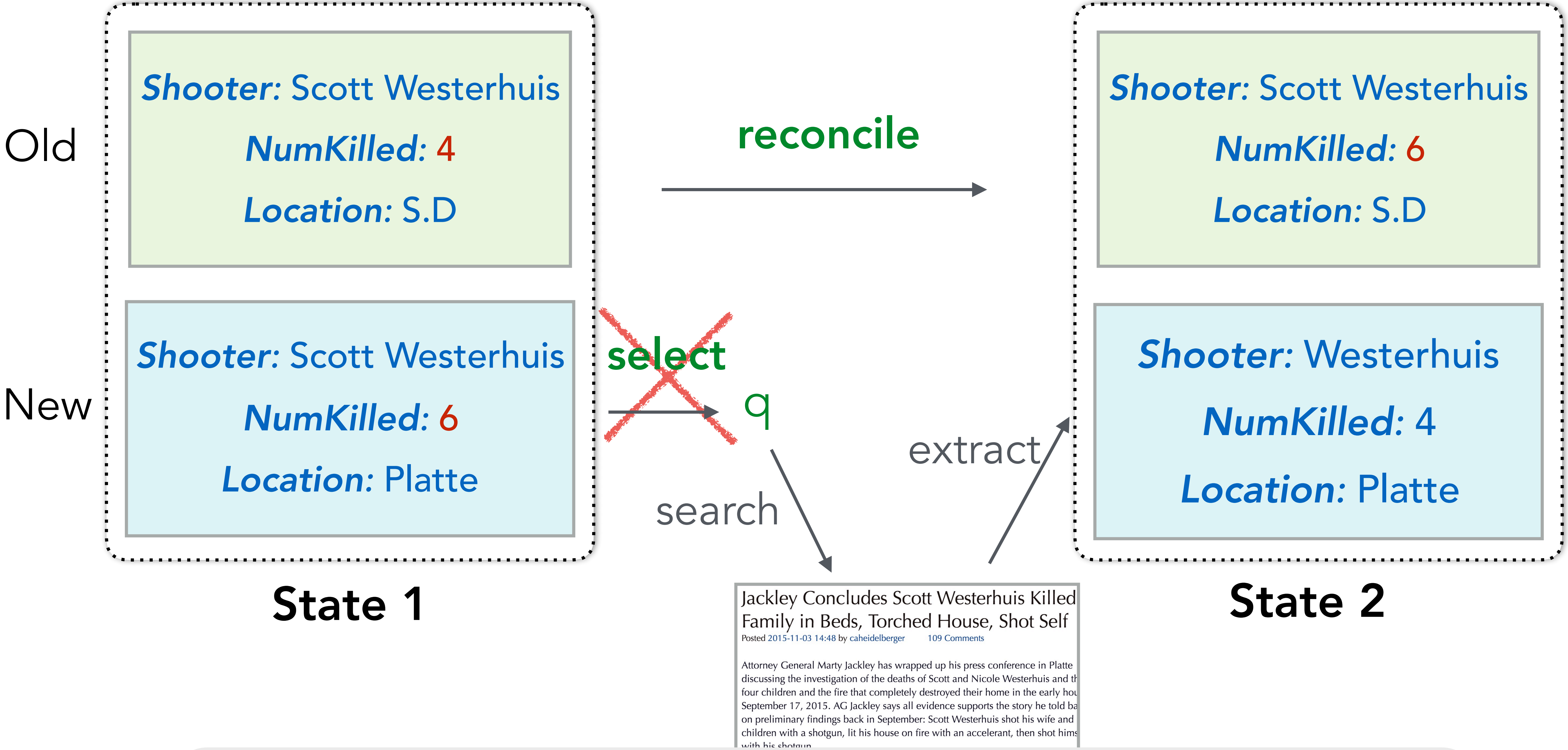
NumKilled: 4

Location: Platte

State 2

Both reconciliation and querying

RL-Basic



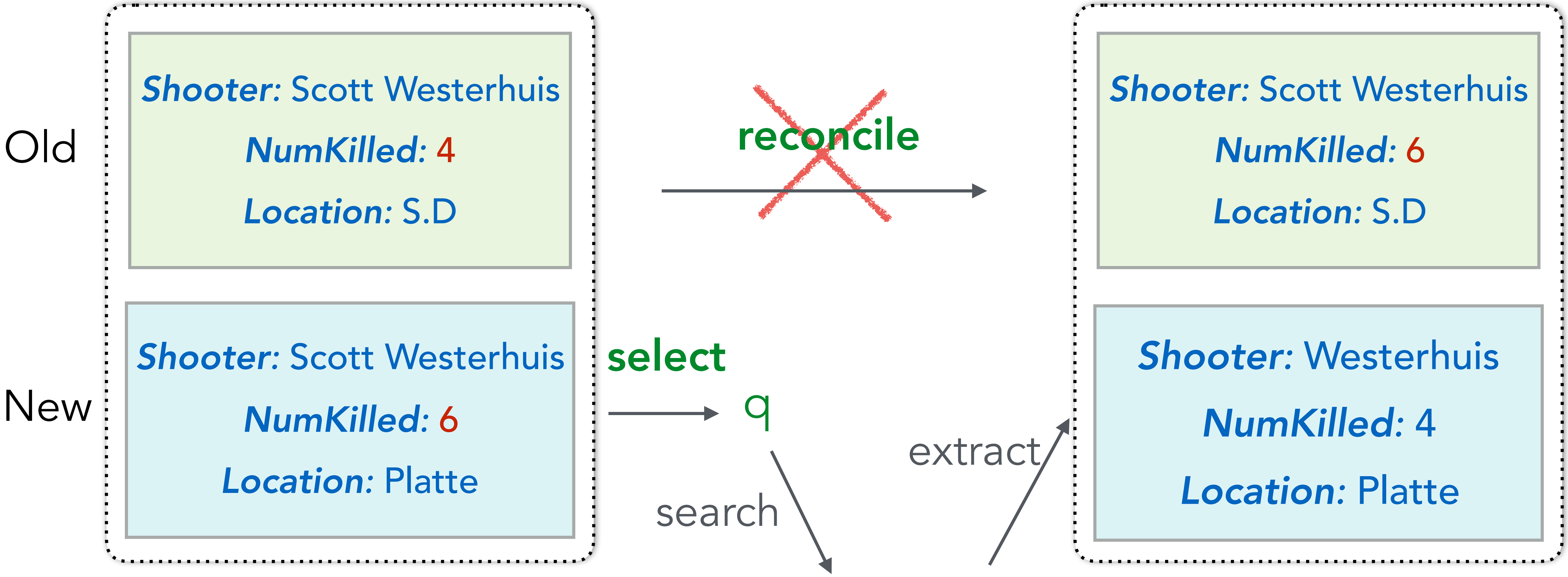
Jackley Concludes Scott Westerhuis Killed Family in Beds, Torched House, Shot Self

Posted 2015-11-03 14:48 by caheidelberger 109 Comments

Attorney General Marty Jackley has wrapped up his press conference in Platte discussing the investigation of the deaths of Scott and Nicole Westerhuis and th four children and the fire that completely destroyed their home in the early hou September 17, 2015. AG Jackley says all evidence supports the story he told ba on preliminary findings back in September: Scott Westerhuis shot his wife and children with a shotgun, lit his house on fire with an accelerant, then shot hims with his shotgun.

Documents are presented in round robin order from different query lists

RL-Query



Old

Shooter: Scott Westerhuis
NumKilled: 4
Location: S.D

New

Shooter: Scott Westerhuis
NumKilled: 6
Location: Platte

~~reconcile~~

select

q

search

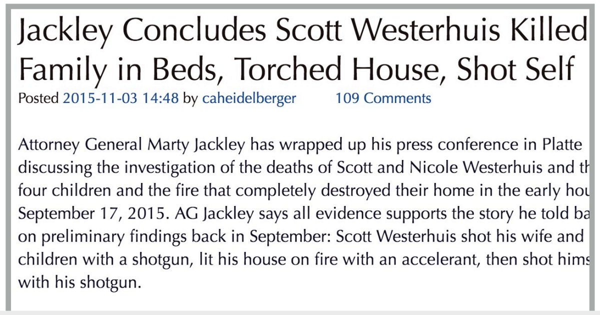
extract

Shooter: Scott Westerhuis
NumKilled: 6
Location: S.D

Shooter: Westerhuis
NumKilled: 4
Location: Platte

State 1

State 2



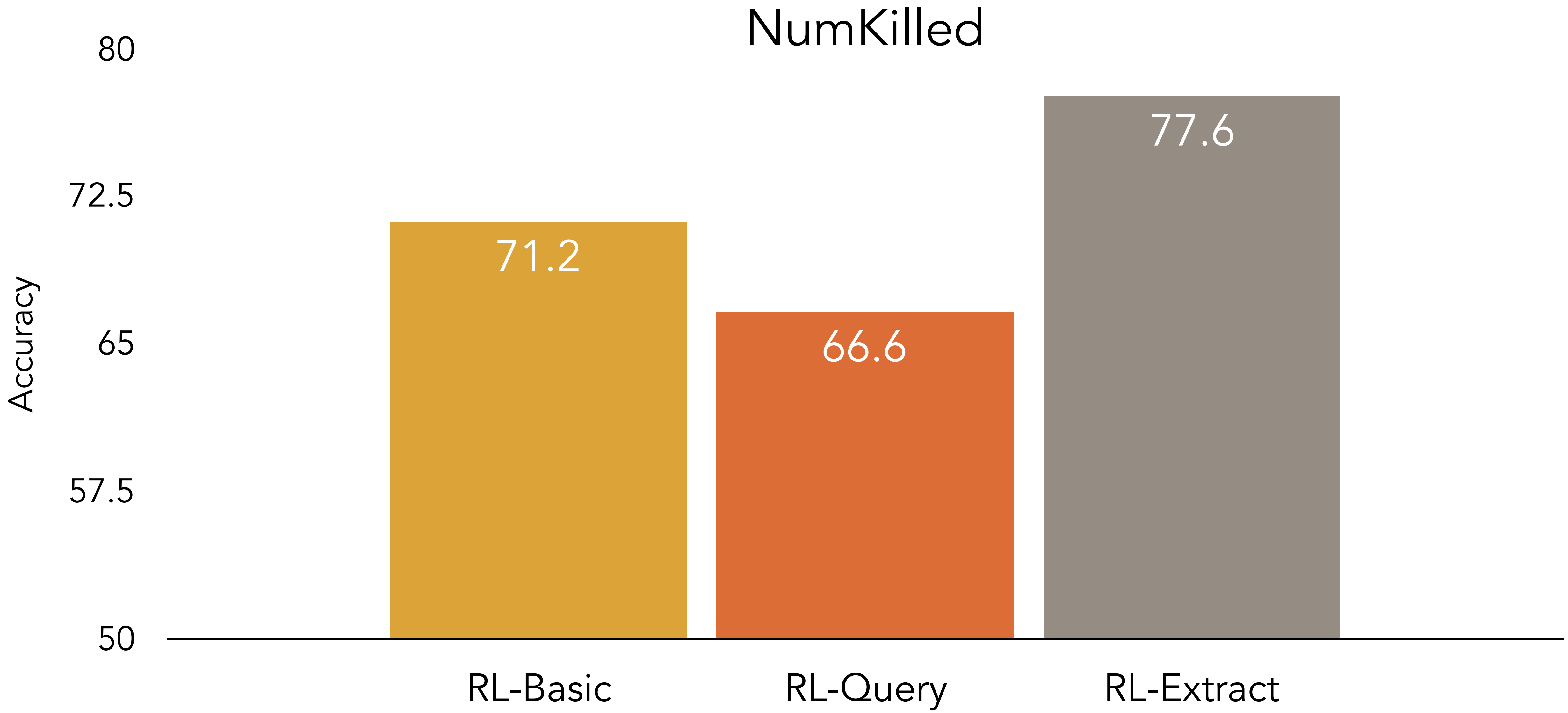
Jackley Concludes Scott Westerhuis Killed Family in Beds, Torched House, Shot Self

Posted 2015-11-03 14:48 by caheidelberger 109 Comments

Attorney General Marty Jackley has wrapped up his press conference in Platte
discussing the investigation of the deaths of Scott and Nicole Westerhuis and th
four children and the fire that completely destroyed their home in the early hou
September 17, 2015. AG Jackley says all evidence supports the story he told ba
on preliminary findings back in September: Scott Westerhuis shot his wife and
children with a shotgun, lit his house on fire with an accelerant, then shot hims
with his shotgun.

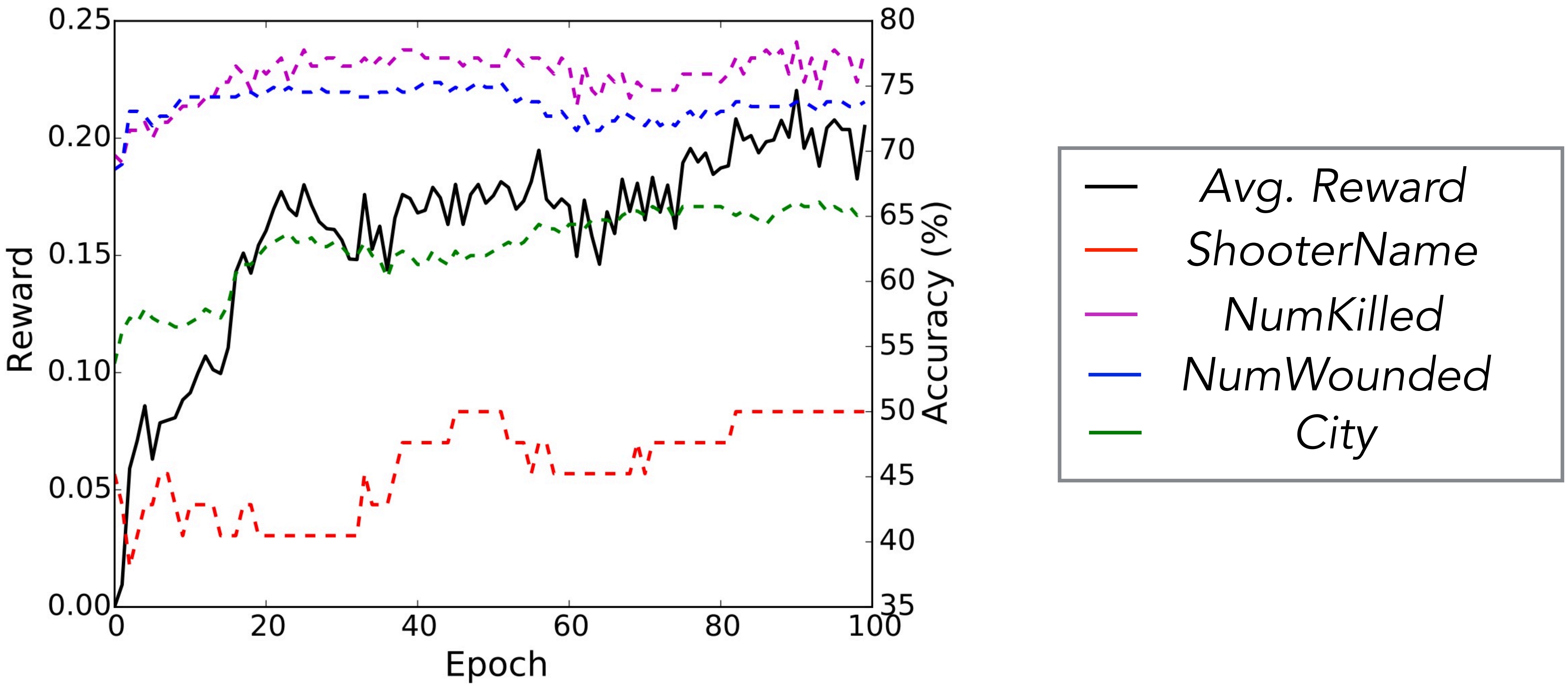
Reconciliation is confidence-based

RL Models



Both reconciliation and querying are important and inter-linked

Evolution of Test Accuracy



Agent learns to balance all entity choices simultaneously

Examples

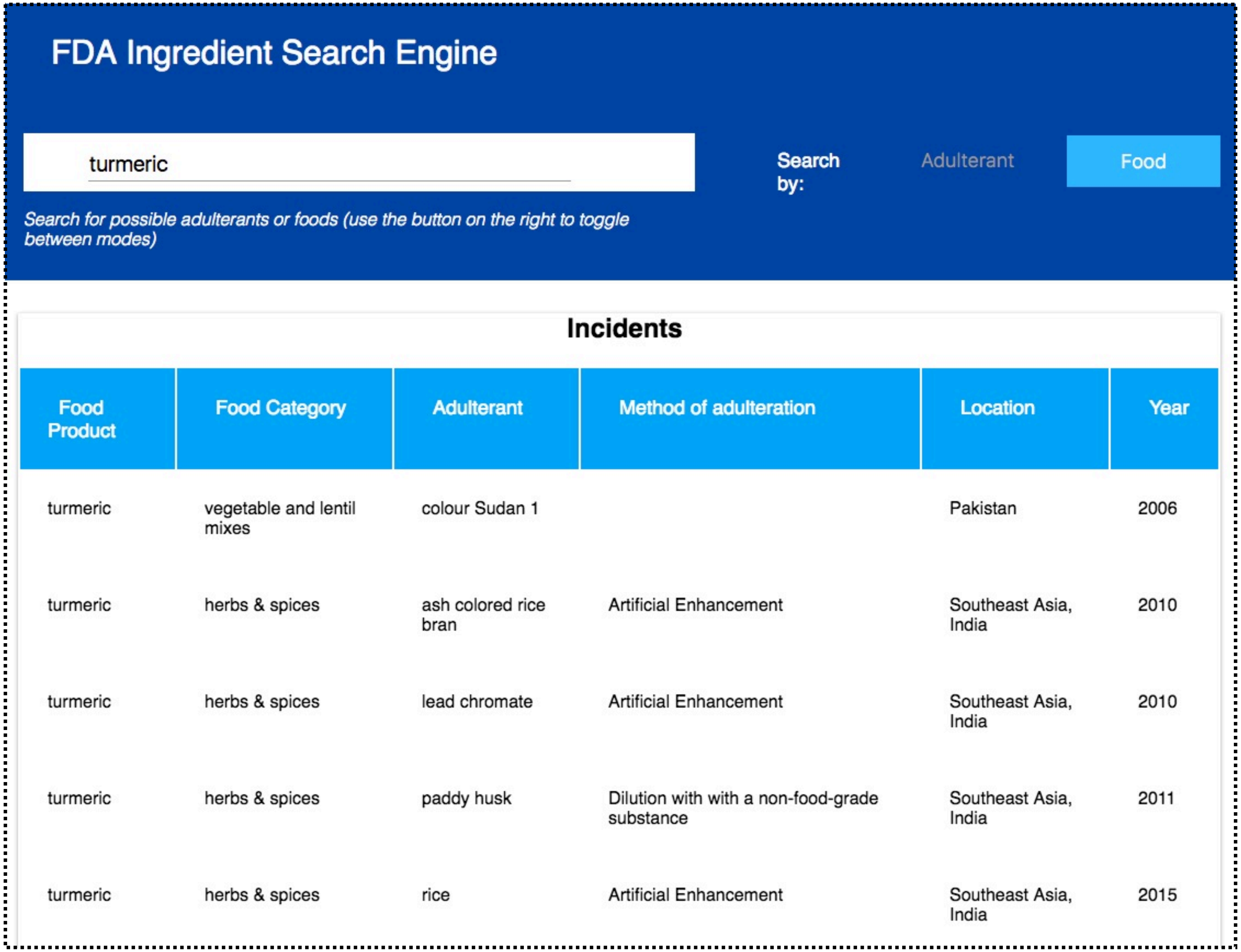
	Text	**Shooter Name**
Basic Extractor	A source tells Channel 2 Action News that Thomas Lee has been arrested in Mississippi ... Sgt . **Stewart** Smith, with the Troup County Sheriff's office, said.	Stewart
RL-Extract	**Lee** is accused of killing his wife, Christie; ...	Lee

Examples

	Text	**NumKilled**
Basic Extractor	Shooting leaves 25 year old Pittsfield man dead , 4 injured	**0**
RL-Extract	**One** man is dead after a shooting Saturday night at the intersection of Dewey Avenue and Linden Street.	**1**

Our system finds alternative sources of information for reliable extraction

Adulteration Detection



FDA Ingredient Search Engine

turmeric

Search for possible adulterants or foods (use the button on the right to toggle between modes)

Incidents

Food Product	Food Category	Adulterant	Method of adulteration	Location	Year
turmeric	vegetable and lentil mixes	colour Sudan 1		Pakistan	2006
turmeric	herbs & spices	ash colored rice bran	Artificial Enhancement	Southeast Asia, India	2010
turmeric	herbs & spices	lead chromate	Artificial Enhancement	Southeast Asia, India	2010
turmeric	herbs & spices	paddy husk	Dilution with with a non-food-grade substance	Southeast Asia, India	2011
turmeric	herbs & spices	rice	Artificial Enhancement	Southeast Asia, India	2015

Conclusion

- Alternative paradigm to improve Information Extraction, especially for low-resource domains.
- Use of Reinforcement Learning to find and incorporate external information.

Code and data available at:
http://people.csail.mit.edu/karthikn/rl-ie/