

Compositional Questions Do Not Necessitate Multi-hop Reasoning

Sewon Min*, Eric Wallace*, Sameer Singh, Matt Gardner,
Hannaneh Hajishirzi, Luke Zettlemoyer

University of Washington



Allen Institute for Artificial Intelligence



UC Irvine





Sewon Min
UW



Eric Wallace
AI2



Sameer Singh
UCI



Matt Gardner
AI2



Hannanah Hajishirzi
UW, AI2



Luke Zettlemoyer
UW

Multi-hop Question Answering

Q: What is the former name of the Bonobo Apes?



Single-hop QA

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?



Multi-hop QA

Multi-hop Question Answering

Compositional

Q: What is the former name of the Bonobo Apes?



Single-hop QA

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?



Multi-hop QA

Multi-hop Question Answering

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?

Multi-hop Question Answering

Q: What is the former name of the animal whose **habitat** the Reserve Naturelle Lomako Yokokala was established to protect?



P1: The Lomako Forest Reserve was established to protect the **habitat** of the **Bonobo apes**

Multi-hop Question Answering

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?



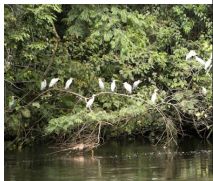
P1: The Lomako Forest Reserve was established to protect the habitat of the Bonobo apes



P2: The bonobo, formerly called the pygmy chimpanzee, is an endangered great ape ...

Multi-hop Question Answering

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Y



**Compositional Questions
== Multi-hop Questions ???**

protect



P2: The bonobo, formerly called the pygmy chimpanzee, is an endangered great ape ...

Research Questions

1. Do compositional questions necessarily require multi-hop?
2. How many are solvable in single-hop?
3. What factors lead to single-hop?
4. Can we turn them into multi-hop?

Research Questions

1. Do compositional questions necessarily require multi-hop?
2. How many are solvable in single-hop?
3. What factors lead to single-hop?
4. Can we turn them into multi-hop?

Distractor Setting

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?

P1: The Lomako Forest Reserve was established to protect the habitat of the Bonobo apes.

P2: The bonobo, formerly called the pygmy chimpanzee is (...)

P3 Nature Reserve of Saint Bartholomew

P4 Grand Cote National Wildlife Refugei

P5 La Tortue, Saint Barthelemy

P6 Reserve Naturelle Lomoko Yokokala

P7 Cereste

P8 Guadeloupe National Park

P9 Ile fourchue

P10 La National Nature Reserve

Distractor Setting

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?

P1: The Lomako Forest Reserve was established to protect the habitat of the Bonobo apes.

P2: The bonobo, formerly called the pygmy chimpanzee is (...)

P3 Nature Reserve of Saint Bartholomew

P4 Grand Cote National Wildlife Refugei

P5 La Tortue, Saint Barthelemy

P6 Reserve Naturelle Lomoko Yokokala

P7 Cereste

P8 Guadeloupe National Park

P9 Ile fourchue

P10 La National Nature Reserve

No other paragraph about an animal
→ Can be Single hop in distractor setting

Full-wiki Setting

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?



WIKIPEDIA
The Free Encyclopedia

Common wood pigeon

From Wikipedia, the free encyclopedia

For other uses, see *Wood pigeon* (disambiguation).

The **common wood pigeon** (*Columba palumbus*) is a large species in the dove and pigeon family. It belongs to the genus *Columba* and, like all pigeons and doves, belongs to the family Columbidae. It is locally known in southeast England as the "coulver".^[2] This name has given rise to several areas known for keeping pigeons to be named after it, such as *Culver Down*.

The genus name *Columba* is the Latin word meaning "pigeon, dove".^[2] whose older etymology comes from the Ancient Greek κόλυμβος (*kolumbos*), "a diver", from κόλυβιδας (*kolumbas*), "dive, plunge headlong, swim".^[2] *Aristophanes*

Common wood pigeon



Perched on a garden fence post in England

Roe deer

From Wikipedia, the free encyclopedia

The **European roe deer** (*Capreolus capreolus*), also known as the **western roe deer**, *chevreuil*, or simply **roe deer** or **roe**, is a species of **deer**. The male of the species is sometimes referred to as a **roebuck**. The roe deer is relatively small, reddish and grey-brown, and well-adapted to cold environments. The species is widespread in Europe, from the **Mediterranean** to **Scandinavia**, from **Scotland** to the **Caucasus**, and east to northern Iran and Iraq. It is distinct from the somewhat larger **Siberian roe deer**.

Asian golden cat

From Wikipedia, the free encyclopedia

Not to be confused with *African golden cat* or *Bay cat*.

It (*Catopuma temminckii*) is a medium-sized cat of the northeastern Indian subcontinent. It has been listed as **Near Threatened** on the IUCN Red List since 2008, and is threatened by **hunting** and **habitat loss**, since Southeast Asian forests are among the world's fastest regional deforestation.^[1] Its scientific name honours the Dutch naturalist **Jacob Temminck**. It is also called **Asiatic golden cat**.^[2]

Asian golden cat



Roe deer



Male (buck) and female (doe) roe deer

Full-wiki Setting

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?



WIKIPEDIA
The Free Encyclopedia

Common wood pigeon

From Wikipedia, the free encyclopedia

For other uses, see Wood pigeon (disambiguation).

The **common wood pigeon** (*Columba palumbus*) is a large species in the dove and pigeon family. It belongs to the genus *Columba* and, like all pigeons and doves, belongs to the family Columbidae. It is locally known in southeast England as the "coulver".^[2] This name has given rise to several areas known for keeping pigeons to be named after it, such as *Culver Down*.

The genus name *Columba* is the Latin word meaning "pigeon, dove";^[2] whose older etymology comes from the *Ancient Greek* *κόλυμβος* (*kolumbos*), "a diver", from *κόλυσσιν* (*kolumbas*), "dive, plunge headlong, swim".^[2] *Aristophanes*

Common wood pigeon




Perched on a garden fence post in England

Roe deer

From Wikipedia, the free encyclopedia

The **European roe deer** (*Capreolus capreolus*), also known as the **western roe deer**, *chevreuil*, or simply **roe deer** or **roe**, is a species of **deer**. The male of the species is sometimes referred to as a **roebuck**. The roe deer is relatively small, reddish and grey-brown, and well-adapted to cold environments. The species is widespread in Europe, from the *Mediterranean* to *Scandinavia*, from *Scotland* to the *Caucasus*, and east to northern Iran and Iraq. It is distinct from the somewhat larger *Siberian roe deer*.

Roe deer



Male (buck) and female (doe) roe deer

Asian golden cat

From Wikipedia, the free encyclopedia

Not to be confused with African golden cat or Bay cat.

It (*Catopuma temminckii*) is a medium-sized feline in the northeastern Indian subcontinent. It has been listed as *Near Threatened* since 2008, and is threatened by *hunting* and *habitat loss*, since Southeast Asian forests are undergoing the fastest regional deforestation.^[1]

Its scientific name honours the Dutch naturalist *Jacob Temminck*. It is also called **Asiatic golden cat**.^[2]

Asian golden cat



Other articles about animals with former names
→ Multi-hop in full-wiki setting

Full-wiki Setting

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomak

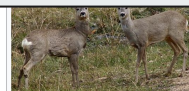
Question can be multi-hop or single-hop depending on the evidence context

WIKIPEDIA
The Free Encyclopedia

[Greek κολυμβός \(kolumbos\), "a diver", from κολυμβάω \(kolumbao\), "dive, plunge headlong, swim" ^{\[2\]} Aristophanes](#)

Perched on a garden fence post
England

species is sometimes referred to as a **roe** **buck**. The roe deer is relatively small, reddish and grey-brown, and well-adapted to cold environments. The species is widespread in Europe, from the Mediterranean to Scandinavia, from Scotland to the Caucasus, and east to northern Iran and Iraq. It is distinct from the somewhat larger [Siberian roe deer](#).



Male (buck) and female (doe) roe deer

[Jacob Temminck](#). It is also called **Asiatic golden cat** ^[2]



Other articles about animals with former names
→ Multi-hop in full-wiki setting

Research Questions

1. Do compositional questions necessarily require multi-hop?
2. How many are solvable in single-hop?
3. What factors lead to single-hop?
4. Can we turn them into multi-hop?

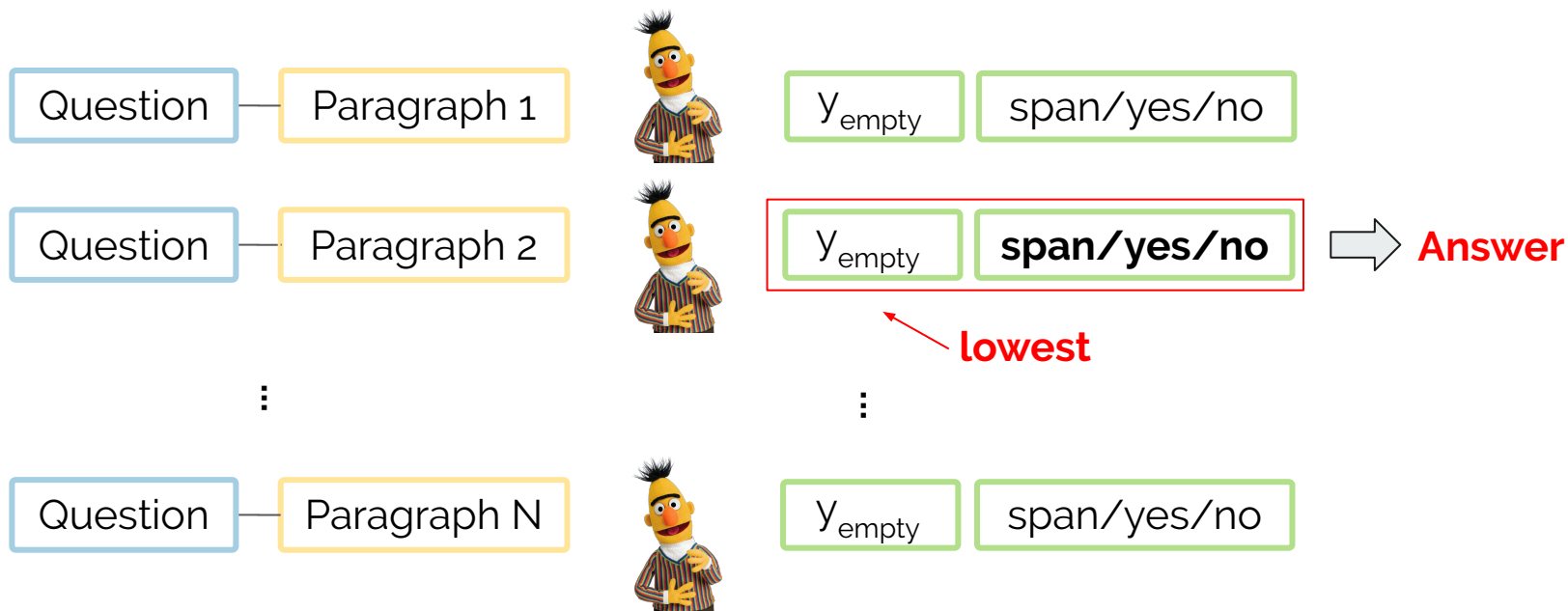
Single-hop QA Model



Single-hop QA Model



Single-hop QA Model



Single-hop QA Result

	Distractor	Full-Wiki
Single-hop QA Model *	67.1	38.4
BiDAF * (Seo et al 17)	58.3	34.4
BiDAF (Seo et al 17)	59.0	32.9
QFE (Nishida et al 19)	68.1	38.1
DecompRC (Min et al 19)	69.6	40.7
DFGN (Xiao et al 19)	69.7	
Cognitive Graph QA (Ding et al 19)	-	48.9

* Official validation set; others are hidden test set
Published models only

Single-hop QA Result

	Distractor	Full-Wiki
Single-hop QA Model *	67.1	38.4
BiDAF * (Seo et al 17)	58.3	34.4
BiDAF (Seo et al 17)	59.0	32.9
QFE (Nishida et al 19)	68.1	38.1
DecompRC (Min et al 19)	69.6	40.7
DFGN (Xiao et al 19)	69.7	
Cognitive Graph QA (Ding et al 19)	-	48.9

* Official validation set; others are hidden test set
Published models only

Humans with Multi-hop

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?

P1: The Lomako Forest Reserve was established in 1991 ...

P2: The bonobo, formerly called the pygmy chimpanzee is ...

P3

P4

P5

P6

P7

P8

P9

P10

87.4 F1



Humans with Single-hop

Q: What is the former name of the animal whose habitat the Reserve Naturelle Lomako Yokokala was established to protect?

87.4 F1



82.1 F1

P2: The bonobo, formerly called the pygmy chimpanzee is ...

P3

P4

P5

P6

P7

P8

P9

P10

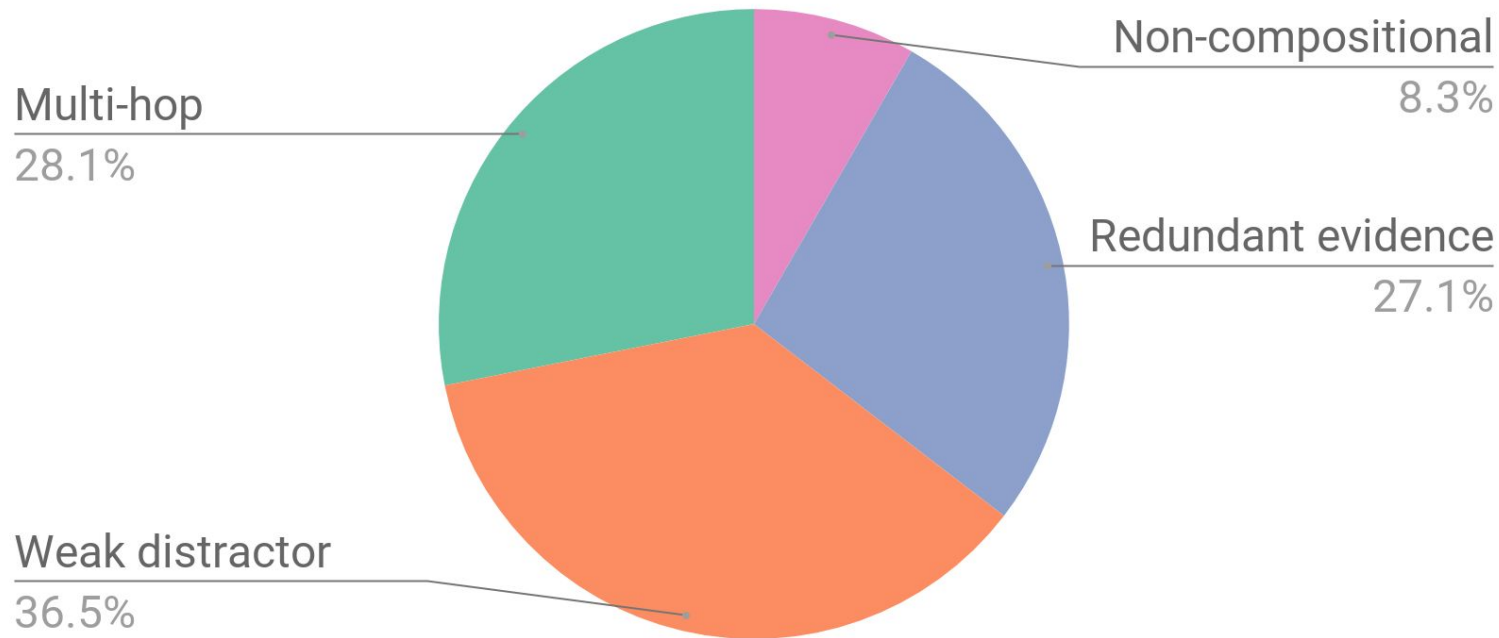


Research Questions

1. Do compositional questions necessarily require multi-hop?
2. How many are solvable in single-hop?
3. What factors lead to single-hop?
4. Can we turn them into multi-hop?

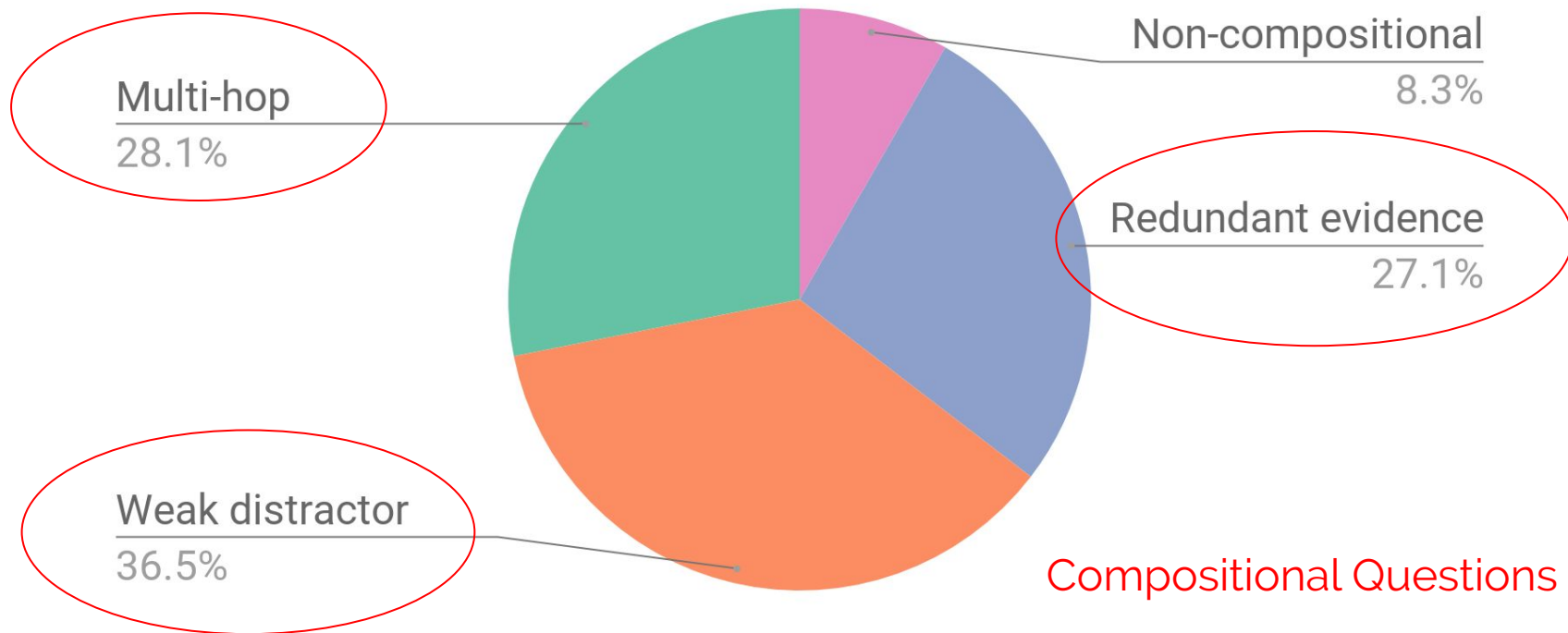
What leads to single-hop?

Categorization of 100 random samples



What leads to single-hop?

Categorization of 100 random samples



What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

Redundant
Evidence

(26%)

Non-compositional
one-hop

(8%)

Q: Ralph Hefferline was a psychology professor at a university that is located in what city?

What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

Redundant

Evidence

(26%)

Non-compositional

one-hop

(8%)

Q: Ralph Hefferline was a psychology professor at a university that is located in what city?

P1: Ralph Hefferline was a psychology professor at Columbia University.

P2: Columbia University is ... in New York City.

P3: Stanley Coren is a psychology professor at the University of British Columbia in Vancouver, British Columbia.

What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

Redundant

Evidence

(26%)

Non-compositional

one-hop

(8%)

Q: Ralph Hefferline was a psychology professor at a university that is located in what city?

P1: Ralph Hefferline was a psychology professor at Columbia University.

P2: Columbia University is ... in New York City.

P3: Stanley Coren is a psychology professor at the University of British Columbia in Vancouver, British Columbia.

What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

Redundant

Evidence

(26%)

Non-compositional

one-hop

(8%)

Q: Ralph Hefferline was a psychology professor at a university that is located in what city?

P1: Ralph

P2: Colu

P3: Stanley Coren is a psychology professor at the University of British Columbia in Vancouver, British Columbia.

Distractor setting



Full-wiki setting



What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

Redundant
Evidence

(26%)

Non-compositional
one-hop

(8%)

Q: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell?

What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

Redundant

Evidence

(26%)

Non-compositional

one-hop

(8%)

Q: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell?

P1: Shirley Temple Black was an American actress, singer, dancer, businesswoman (..), she also served as **Chief of Protocol of the United States**.

What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

Redundant

Evidence

(26%)

Non-compositional

one-hop

(8%)

Q: What government position was held by the woman who portrayed Corliss Archer in the film Kiss and Tell?

P1: Shirley Temple Black was an American actress, singer, dancer, businesswoman (..), she also served as **Chief of Protocol of the United States**.

No other paragraphs are about actresses *with* government positions



What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

Redundant

Evidence

(26%)

Non-compositional

one-hop

(8%)

Q: What government position was held by the woman who portrayed Carliss Archer in the 1950s?

P1: Shirley Temple was a famous actress who appeared in several movies in the 1930s and 1940s. She was also a singer and dancer. She was married to Charles H. Busby in 1942 and they had a son, Charles H. Busby Jr., in 1943. She died of cancer in 1956.

Distractor setting



Full-wiki setting



No other paragraphs are about actresses *with* government positions



What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

**Redundant
Evidence**

(26%)

Non-compositional
one-hop

(8%)

Q: Kaiser Ventures corporation was founded by an American industrialist who became known as the father of modern American shipbuilding ?

What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

**Redundant
Evidence**

(26%)

Non-compositional
one-hop

(8%)

Q: Kaiser Ventures corporation was founded by an American industrialist who became known as the father of modern American shipbuilding?

P1: Kaiser Ventures corporation was founded by Henry J. Kaiser to provide steel plate for the Pacific Coast shipbuilding industry.

Only one founder of Kaiser Ventures; can ignore “American shipbuilding..”



What leads to single-hop?

Multi-hop

(27%)

Weak Distractor

(35%)

Redundant

Evidence

(26%)

Non-compositional

one-hop

(8%)

Q: Kaiser Ventures corporation was founded by an American industrialist who became

P1: Kais
plate fo

Distractor setting



Full-wiki setting



steel

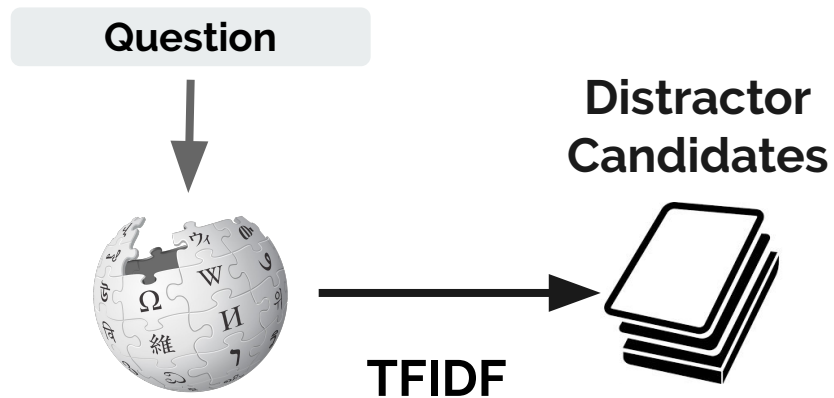
Only one founder of Kaiser Ventures; can ignore "American shipbuilding.."



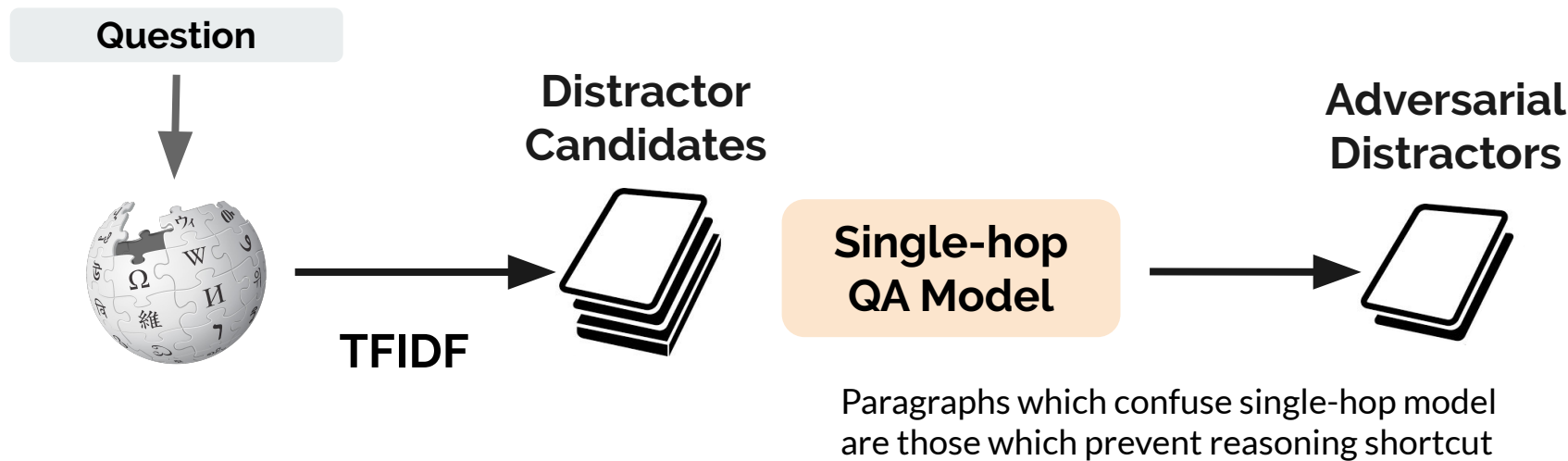
Research Questions

1. Do compositional questions necessarily require multi-hop?
2. How many are solvable in single-hop?
3. What factors lead to single-hop?
4. Can we turn them into multi-hop?

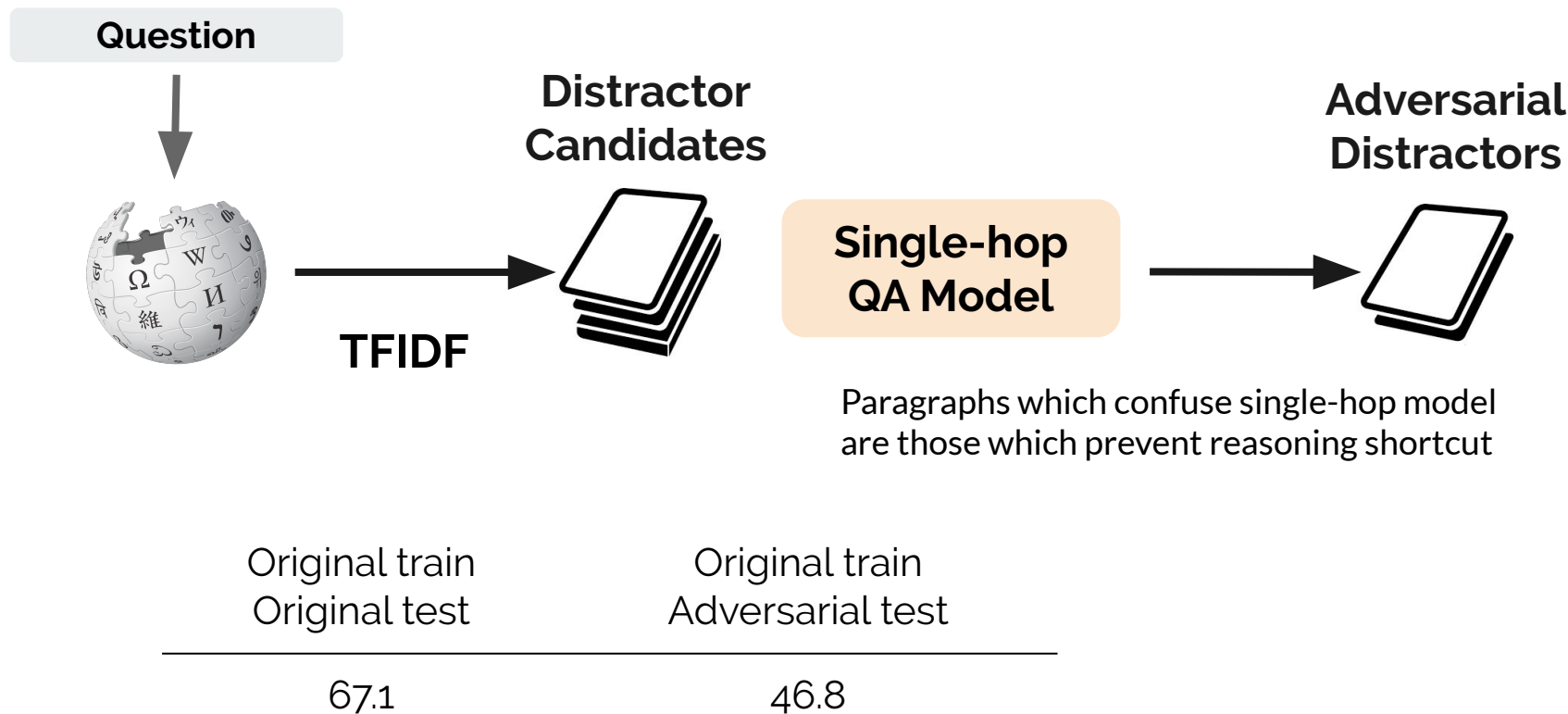
Can We Collect Stronger Distractors?



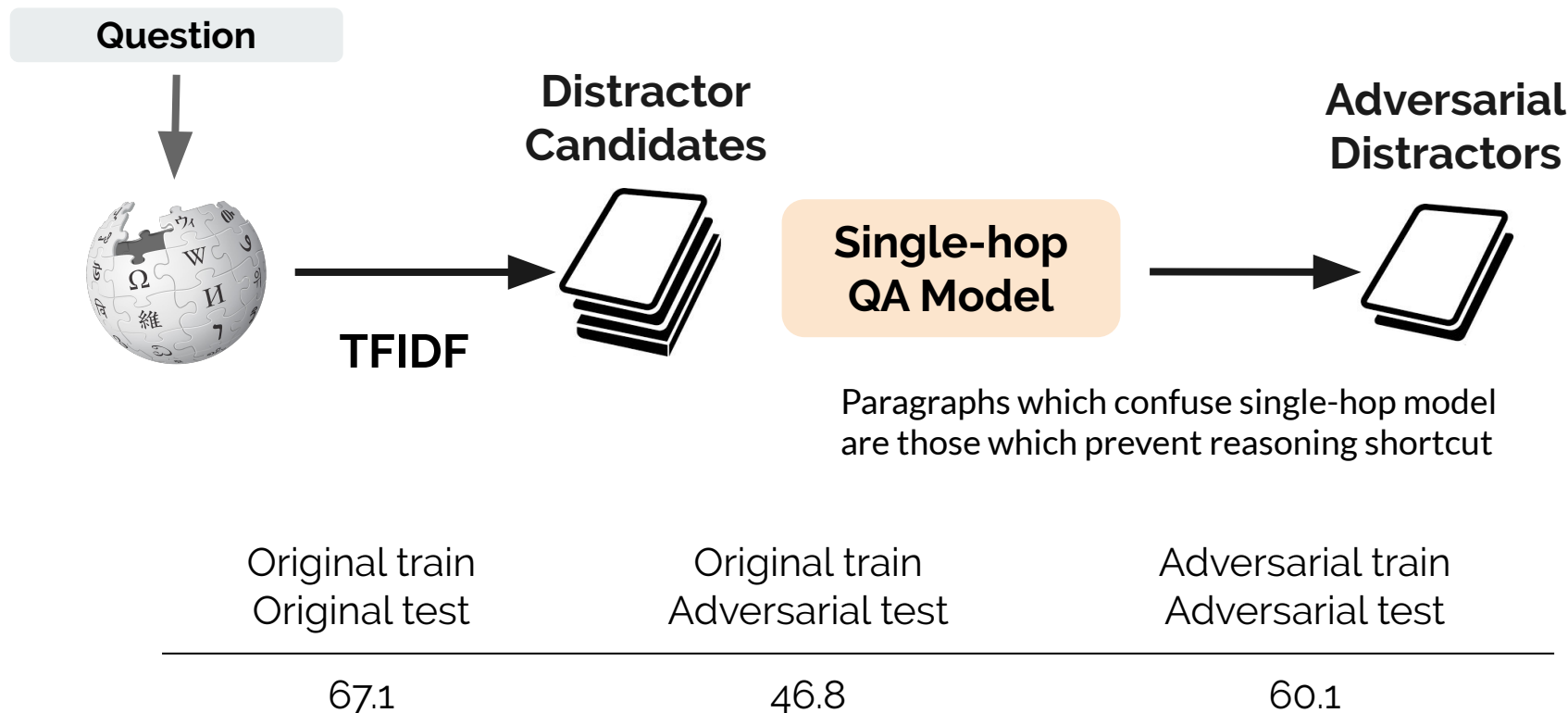
Can We Collect Stronger Distractors?



Can We Collect Stronger Distractors?



Can We Collect Stronger Distractors?



Related Work

Chen & Durrett (NAACL 2019)

- Show that single-hop-only QA model works on par with the state-of-the-art on HotpotQA and WikiHop.

Jiang & Bansal (ACL 2019)

- Reveal reasoning shortcut, and create adversarial examples by editing the contents of the distractors..

Jifan Chen and Greg Durrett. 2019. "Understanding dataset design choices for multi-hop reasoning". NAACL.

Yichen Jiang, Mohit Bansal, 2019. "Avoiding Reasoning Shortcuts: Adversarial Evaluation, Training, and Model Development for Multi-Hop QA". ACL.

Related Work

Ch

Our work is unique in that...

- Reveal that multi-hop depends on the evidence context.
- Categorize different factors.
- Focus on real, retrieval-based adversarial distractors, and demonstrate standard IR system fails.

J

Jifan Chen and Greg Durrett. 2019. "Understanding dataset design choices for multi-hop reasoning". NAACL.

Yichen Jiang, Mohit Bansal, 2019. "Avoiding Reasoning Shortcuts: Adversarial Evaluation, Training, and Model Development for Multi-Hop QA". ACL.

Takeaways

Open-domain multi-hop QA

- Many single-hop Qs in distractor setting are multi-hop in open-domain setting
- Single-hop QA model is still far from SOTA

Collecting strong distractors for multi-hop

- No obvious way to collect strong distractors
 - May be as difficult as solving open-domain multi-hop QA

Thank you for listening!

Code: <https://github.com/shmsw25/single-hop-rc>

Paper: <https://arxiv.org/abs/1906.02900>

