Knowledge representation: Evidence from the cognitive science of counterfactual reasoning

Ruth Byrne

Trinity College Dublin, University of Dublin, Ireland

Cognitive Aspects of Knowledge Representation, IJCAI Workshop 23 July 2022

1

Discoveries about human knowledge representation as a resource for developments in AI

Cognitive science research on counterfactuals to illustrate

2

Counterfactual thoughts

We often imagine how things could have turned out differently, "if only…" usually after bad outcomes

Explanations

Counterfactual possibilities help us to explain the past, we can use them to work out causal relations

Intentions

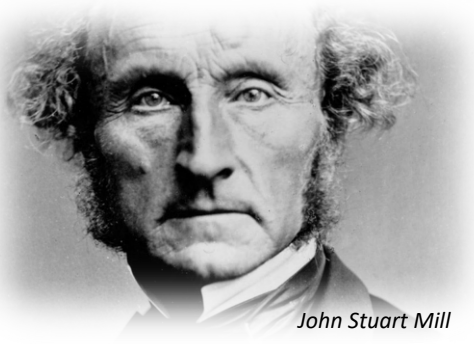
They help us to prepare for the future, we can learn from mistakes, form intentions, make decisions

Kahneman & Tversky, 1982; Ferrante et al., 2013; Roese & Epstude, 2017; Spellman & Mandel, 1999
Byrne 2016 Annual Review of Psychology

3

Counterfactual explanations

David Hume



John Stuart Mill

Counterfactuals and Causes
Two sides of one coin?

If A hadn't happened, B wouldn't have happened = A caused B

Hume, 1739/1978; Mill, 1843/1967

4

Mark Keane
University College Dublin

Greta Warren

Xinyue Dai

Lenart Celar

Counterfactual explanations for decisions of AI systems in XAI

Counterfactual vs causal explanations, pre-factual explanations
Counterfactual explanations for familiar vs unfamiliar domains

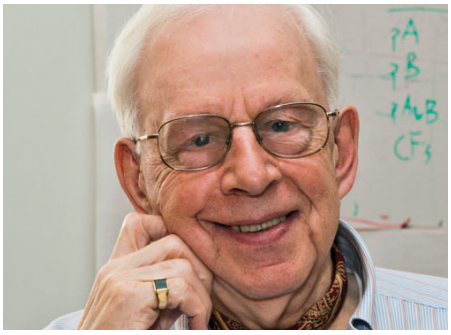
Byrne, 2019, IJCAI; Keane, Kenny, Delaney & Smyth, 2021, IJCAI
Dai, Keane, Shalloo, Ruelle, & Byrne, 2022, AIES

5

Discoveries about human knowledge representation

1. People construct iconic mental simulations of possibilities;
2. They can construct multiple representations;
3. They track the epistemic status of their representations;
4. Mental representations are iconic, but can include symbolic information;
5. People prioritize knowledge of background conditions over presuppositions.

6



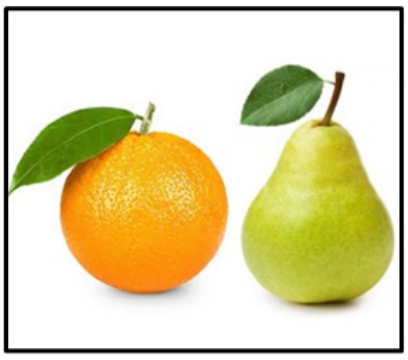
Phil Johnson-Laird
Princeton University

1. People construct iconic mental models to mentally simulate possibilities

Byrne & Johnson-Laird, 2009, Trends in Cognitive Science
Johnson-Laird & Byrne, 2002, Psychological Review

7

If there are oranges, there are pears



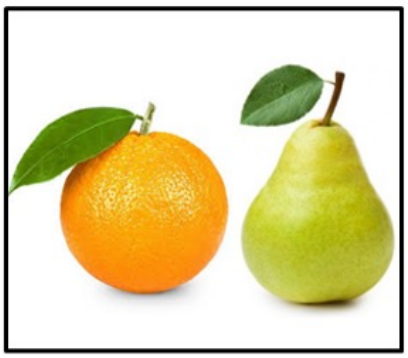
oranges	pears
no oranges	no pears
no oranges	pears

Possibilities

Byrne & Johnson-Laird, 2009, Trends in Cognitive Science
Johnson-Laird & Byrne, 2002, Psychological Review

8

If there are oranges, there are pears



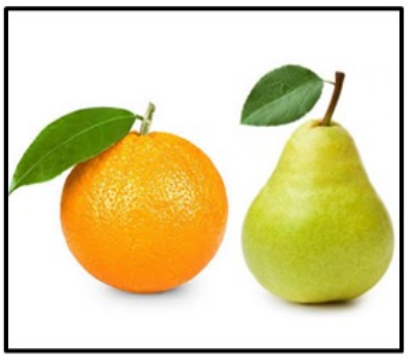
oranges	pears
no oranges	no pears
no oranges	pears
~~oranges~~	~~no pears~~

Principle of truth

Byrne & Johnson-Laird, 2009, Trends in Cognitive Science
Johnson-Laird & Byrne, 2002, Psychological Review

9

If there are oranges, there are pears



oranges	pears
…	

Principle of parsimony

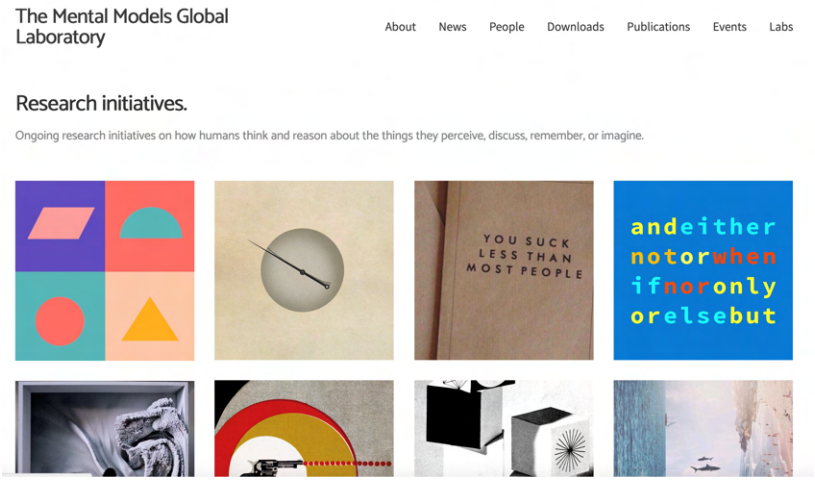
Working memory constraints
Initial, intuitive representation

Byrne & Johnson-Laird, 2009, Trends in Cognitive Science
Johnson-Laird & Byrne, 2002, Psychological Review

10

People construct iconic mental simulations of possibilities

Sangeet Khemlani

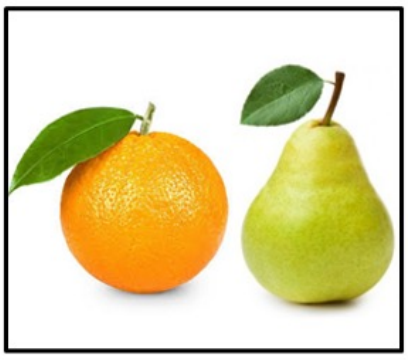


Experimental evidence, computational simulations

Knowledge representation in AI:
how people represent possibilities

11

If there had been oranges, there would have been pears



oranges	pears	*counterfactual*
no-oranges	no-pears	*facts*
…		

Initial, intuitive representation for a counterfactual contains multiple possibilities

Byrne, 2005, *The Rational Imagination*, MIT press

12

Isabel Orenes
UNED Madrid

2. Do people simulate multiple possibilities when they understand a counterfactual?

Orenes, Espino, & Byrne, 2021, Quarterly Journal of Experimental Psychology
Orenes, Garcia-Madruga, Gomez-Vega, Espino, & Byrne, 2019, Frontiers in Psychology

13

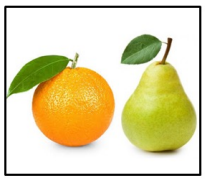
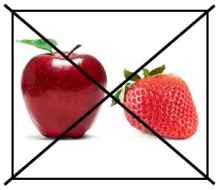
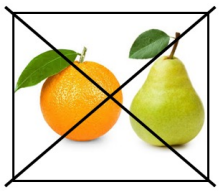
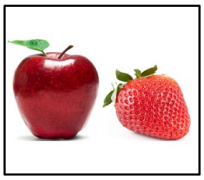
Eye-tracking

Adults listened to short stories on headphones while they looked at displays on screen, cameras recorded their eye movements

Where people look indicates what they are thinking about

14

If there are oranges, there are pears

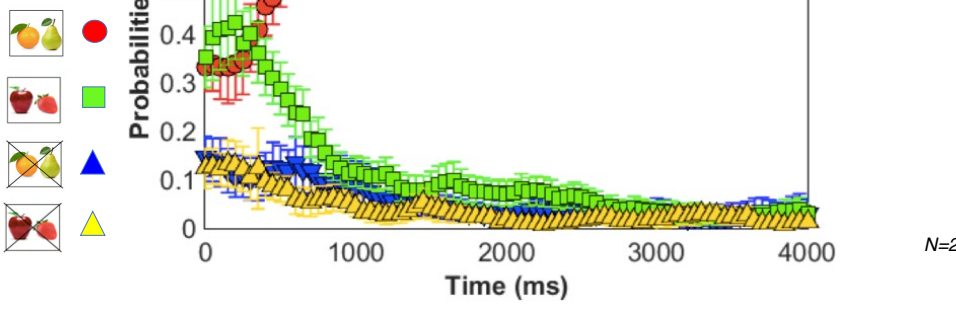





-many different contents
-images or words

Orenes, Garcia-Madruga, Gomez-Vega, Espino & Byrne, 2019, Frontiers in Psychology

15

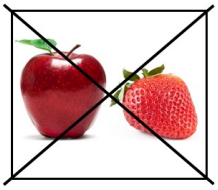
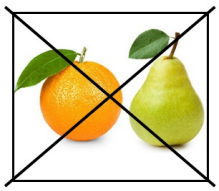
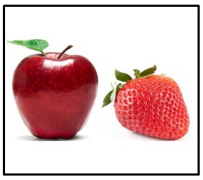
If there are <u>oranges</u> there are pears



N=24 students

16

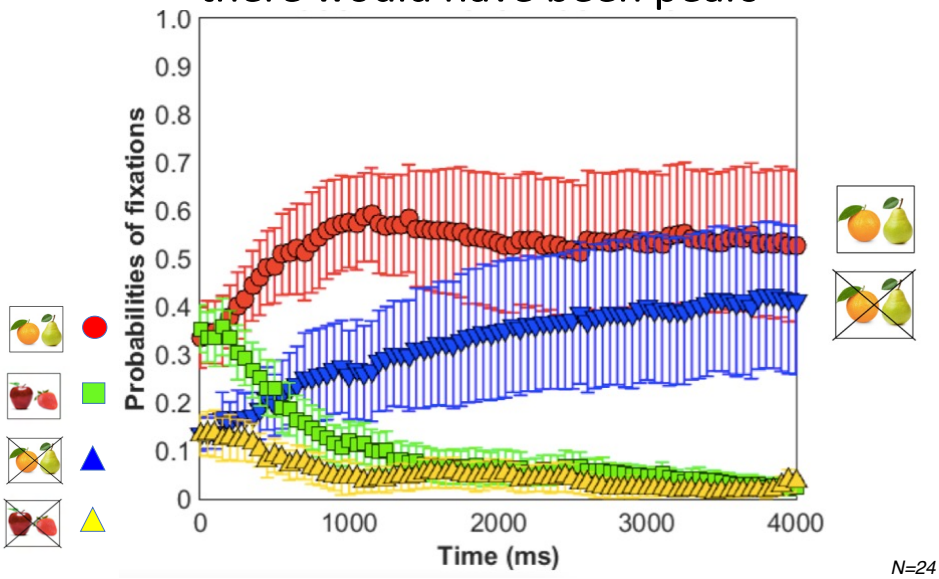
If there had been oranges,
there would have been pears

-many different contents
-images or words

17

If there had been <u>oranges</u>
there would have been pears



N=24 students

Orenes, et al., 2019, Frontiers in Psychology

18

Because she arrived early
she bought roses

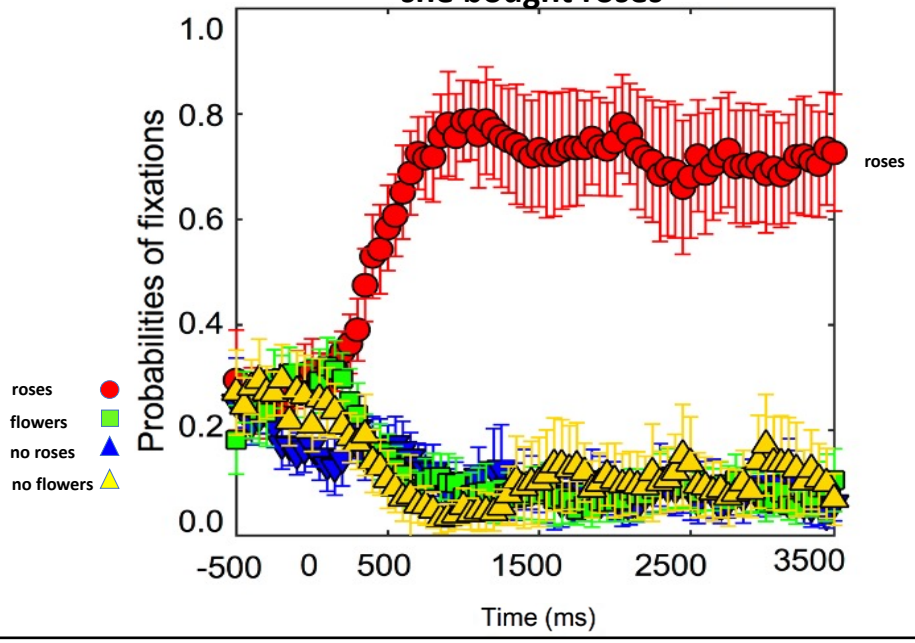
ROSES	FLOWERS
NO ROSES	NO FLOWERS

-many different contents
-words

Orenes, Espino, & Byrne, 2021, Quarterly Journal of Experimental Psychology

19

Because she arrived early
she bought roses



N=71 students

20

If she had arrived early
she would have bought roses

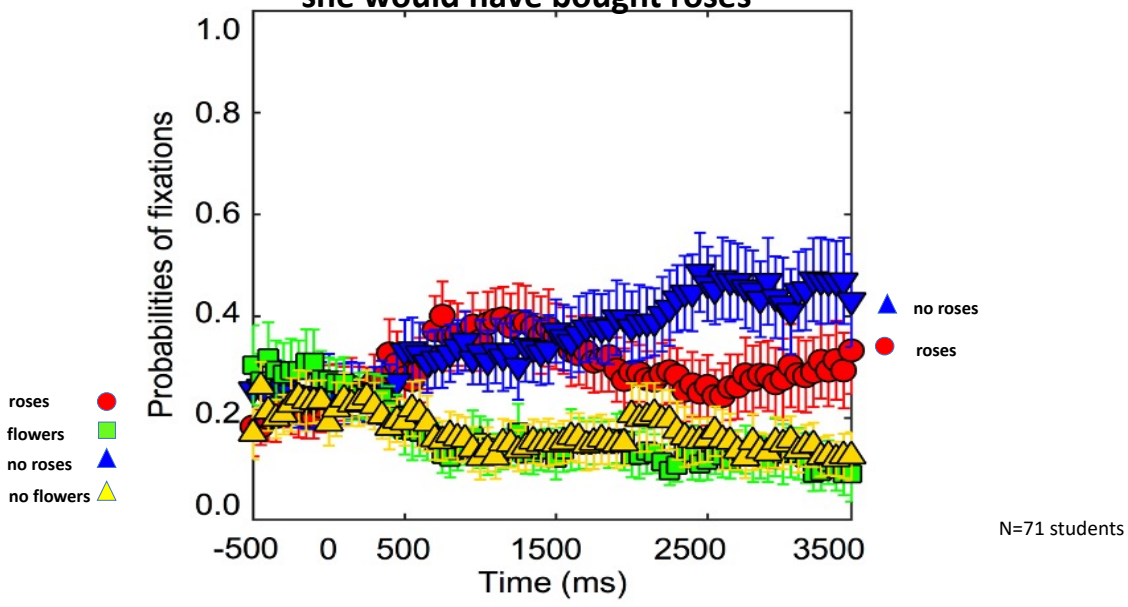
ROSES	**FLOWERS**
NO ROSES	**NO FLOWERS**

-many different contents
-words

Orenes, Espino, & Byrne, 2021, Quarterly Journal of Experimental Psychology

21

If she had arrived early
she would have bought roses



N=71 students

22

People construct multiple representations
when they understand a counterfactual

Knowledge representation in AI:
Counterfactuals provide richer information initially than other sorts of assertions such as causal assertions

23

Orlando Espino
La Laguna University, Tenerife

3. Do people track the epistemic status of representations?

Espino & Byrne, 2021, Journal of Experimental Psychology: Learning, Memory & Cognition

24

Participants read short stories (many different contents), e.g., about tourists visiting a national park, including a counterfactual

Presented one sentence at a time on screen, participants pressed key to see next sentence

Measured latencies to read target sentences corresponding to the counterfactual's conjecture or to the presupposed fact

What people expect they can read quickly

25

The guide tells them

If it had been a good year, there would have been roses

As expected, during the tour they saw

there were no roses < *there were roses*

good	roses	*Counterfactual*
no good	no roses	*Facts*

26

The guide tells them

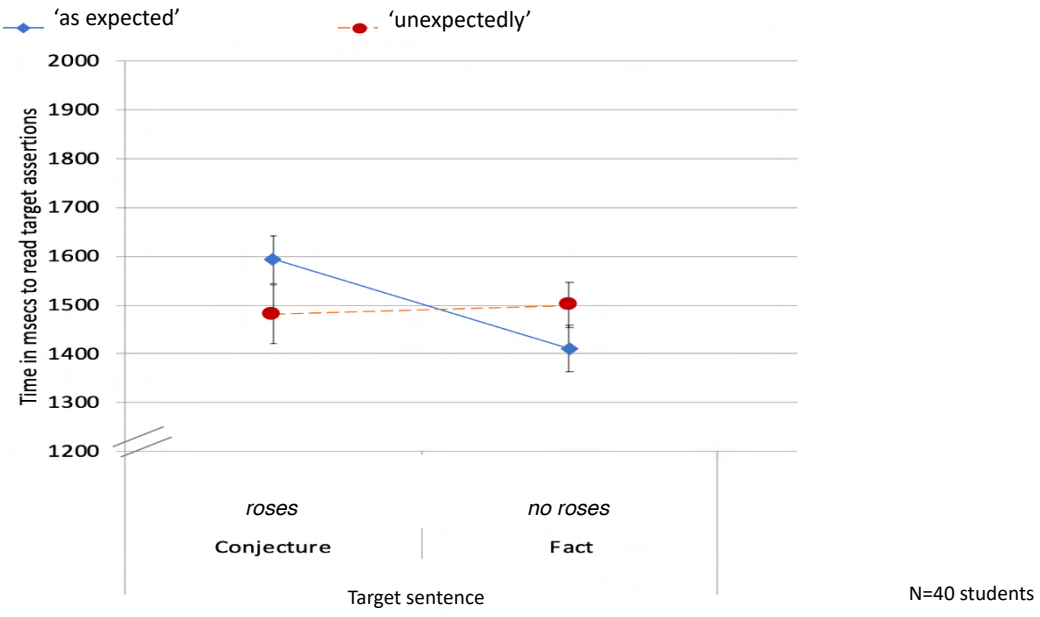
If it had been a good year, there would have been roses

Unexpectedly, *during the tour they saw*

there were no roses = *there were roses*

good	roses	*Counterfactual*
no good	no roses	*Facts*

27



N=40 students

Espino & Byrne, 2021, Journal of Experimental Psychology: Learning, Memory & Cognition

28

People track the epistemic status of their representations

Knowledge representation in AI:
Crucial for people to monitor what is real,
to be able to distinguish a fact about what happened from a hypothetical possibility

Simmons, Garrison & Johnson, 2017; Roese et al., 2008

Espino & Byrne, 2021, Journal of Experimental Psychology: Learning, Memory & Cognition

29

If the flowers had been roses, the trees would have been orange trees

roses	oranges	*Counterfactual*
no roses	no oranges	*Facts*

4. Do mental representations of counterfactuals contain symbols or are they 'embodied' (grounded in perception)?

roses	oranges	*Counterfactual*
poppies	apples	*Facts*

Espino & Byrne, 2018, Cognitive Science

30

The flowers were roses or poppies and the trees were oranges or apples.

If the flowers had been roses, the trees would have been orange trees.

The trees were not orange trees.

What follows?

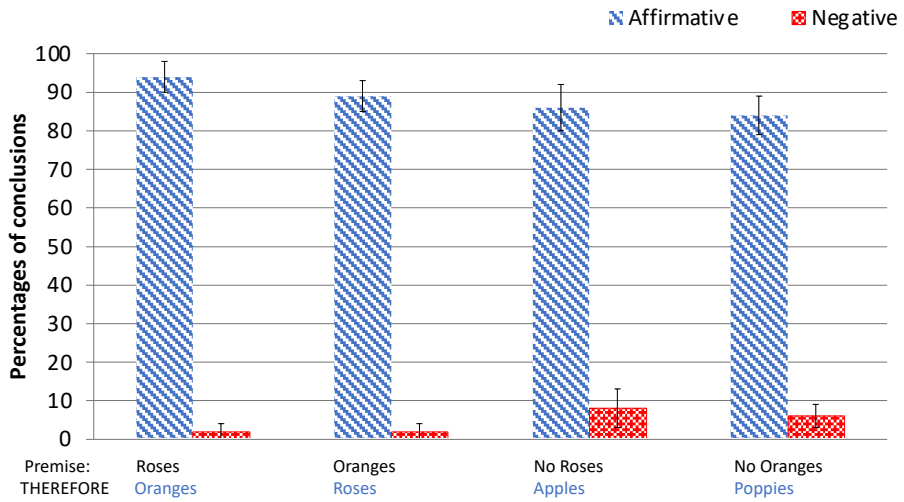
The flowers were poppies.

roses	oranges	*counterfactual*
poppies	apples	*facts*
	…	

31

The flowers were roses or poppies and the trees were oranges or apples.

If the flowers had been roses, the trees would have been orange trees



Espino & Byrne, 2018, Cognitive Science

32

MULTIPLE CONTEXT

The flowers were roses or poppies and the trees were oranges or apples or pears.
If the flowers had been roses, the trees would have been orange trees

The flowers were roses or poppies or lilies and the trees were oranges or apples.
If the flowers had been roses, the trees would have been orange trees

33

MULTIPLE CONTEXT

The flowers were roses or poppies or lilies and the trees were oranges or apples.
If the flowers had been roses, the trees would have been orange trees

roses	orange	*counterfactual*
poppies	apple	*facts*
lilies	apple	
	…	

Alternates
Embodied

Glenberg, et al., 1999;
Kaup, et al., 2006; Mayo, et al., 2004

roses	orange	*counterfactual*
no roses	no orange	*facts*
	…	

Symbols
Pluralist

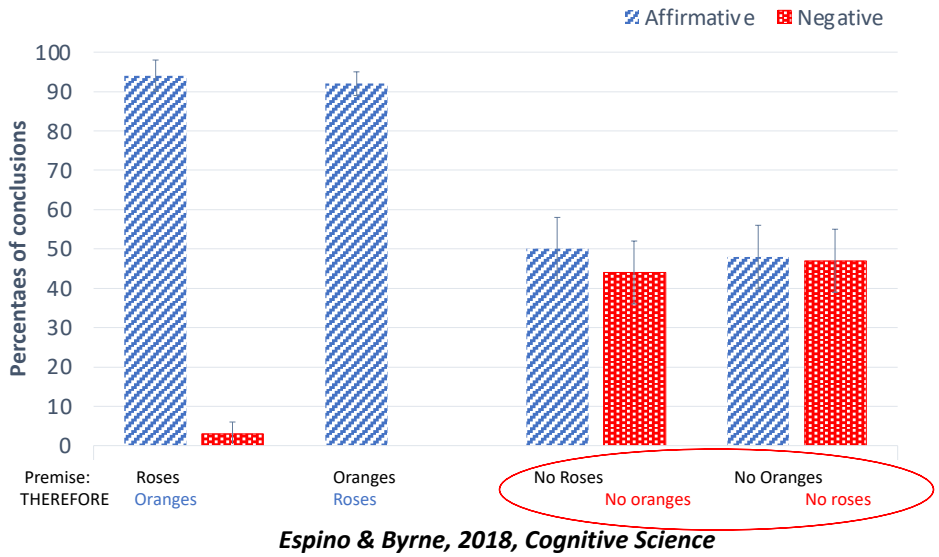
Orenes, Beltan & Santamaria, 2014
Dove, 2009

34

MULTIPLE CONTEXT

The flowers were roses or poppies and the trees were oranges or apples or pears.

If the flowers had been roses, the trees would have been orange trees



Espino & Byrne, 2018, Cognitive Science

35

Mental representations are iconic, but can include symbolic information

Knowledge representation in AI:

People rely on pluralist representations

36

5. How do people incorporate different sorts of knowledge into their mental representations when they make inferences?

Knowledge about presupposed facts and background conditions

Espino & Byrne, 2020, Cognitive Science

37

Factual conditional

If the plants were watered then they grew.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	grow
. . .	

Nothing follows

water	grow
no water	no grow
. . .	

The plants were not watered

Counterfactual conditional

If the plants had been watered then they would have grown.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	grow	*counterfactual fact*
no water	no grow	
. . .		

Counterfactual inference effect

Knowledge about presupposed facts of a counterfactual *increases* inferences

Byrne & Tasso, 1999, Memory & Cognition; Thompson & Byrne, 2002, JEP:LMC

38

Factual conditional

If the plants were watered then they grew.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	grow
	. . .

Nothing follows

water	grow
no water	no grow
	. . .

The plants were not watered

Factual conditional with background conditions

If the plants were watered then they grew.
If the sun shone then they grew
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	sun	grow
water	no sun	no grow
	. . .	

Suppression effect

Knowledge about background conditions of a factual conditional *decreases* inferences

Byrne, 1989, Cognition; Byrne, Espino & Santamaria 1999, Journal of Memory & Language

39

Factual conditional

If the plants were watered then they grew.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	grow
	. . .

water	grow
no water	no grow
	. . .

Factual conditional with background conditions

If the plants were watered then they grew.
If the sun shone then they grew
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	sun	grow
water	no sun	no grow
	. . .	

Knowledge about background conditions decreases inferences

Counterfactual conditional

If the plants had been watered then they would have grown.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	grow	*counterfactual*
no water	no grow	*fact*
	. . .	

Knowledge about presupposed facts of a counterfactual increases inferences

40

Factual conditional

If the plants were watered then they grew.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	grow
	. . .

water	grow
no water	no grow
	. . .

Factual conditional with background conditions

If the plants were watered then they grew.
If the sun shone then they grew
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	sun	grow
water	no sun	no grow
	. . .	

Counterfactual conditional

If the plants had been watered then they would have grown.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	grow	*counterfactual*
no water	no grow	*fact*
	. . .	

Counterfactual conditional with background conditions

?

41

Factual conditional

If the plants were watered then they grew.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	grow
	. . .

water	grow
no water	no grow
	. . .

Factual conditional with background conditions

If the plants were watered then they grew.
If the sun shone then they grew
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	sun	grow
water	no sun	no grow
	. . .	

Counterfactual conditional

If the plants had been watered then they would have grown.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

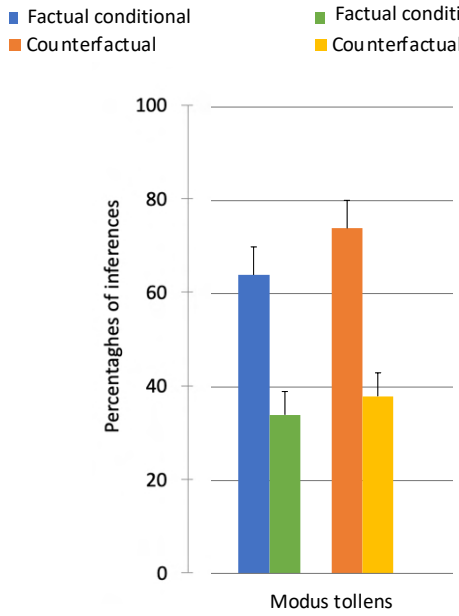
water	grow	*counterfactual*
no water	no grow	*fact*
	. . .	

Counterfactual conditional with background conditions

If the plants had been watered then they would have grown.
If the sun had shone then they would have grown.
The plants did not grow.
Therefore:
(a) The plants were watered.
(b) The plants were not watered.
(c) The plants may or may not have been watered.

water	sun	grow	*counterfactual*
water	no sun	no grow	*fact*
		. . .	

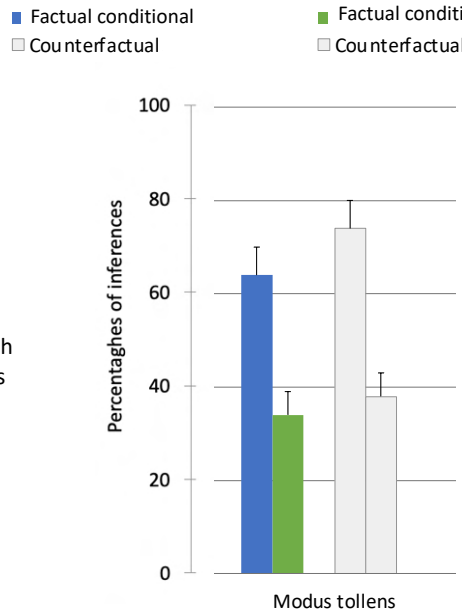
42



N=120, 4 groups
all four inferences,
several contents

Espino & Byrne, 2020, Cognitive Science

43



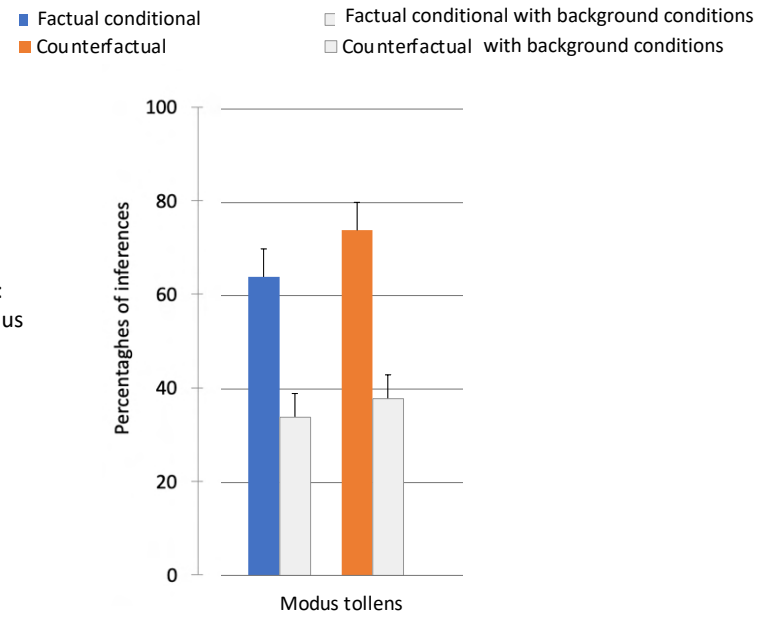
Suppression effect:
Fewer inferences for
factual conditional with
background conditions

N=120, 4 groups
all four inferences,
several contents

Espino & Byrne, 2020, Cognitive Science

44

Counterfactual effect: Somewhat more modus tollens inferences for counterfactual than factual conditional

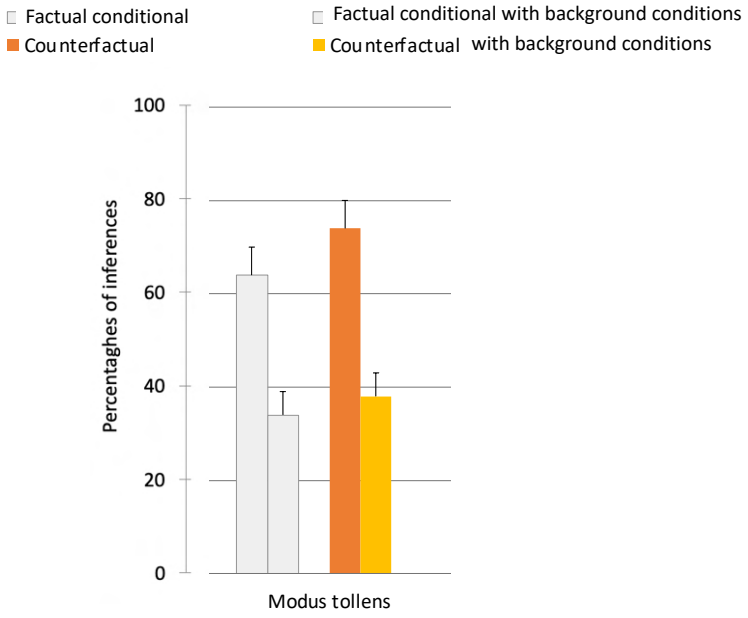


N=120, 4 groups
all four inferences,
several contents

Espino & Byrne, 2020, Cognitive Science

45

Suppression effect for counterfactuals with background conditions



N=120, 4 groups
all four inferences,
several contents

Espino & Byrne, 2020, Cognitive Science

46

People prioritize knowledge of background conditions over presuppositions

Knowledge representation in AI:
Interaction of different sorts of knowledge

47

1. People construct iconic mental simulations of possibilities;
2. They can construct multiple representations;
3. They track the epistemic status of their representations;
4. Mental representations are iconic, but can include symbolic information;
5. People prioritize knowledge of background conditions over presuppositions.

Discoveries about human knowledge representation
as a resource for developments in AI

https://reasoningandimagination.com/ *@ruthmjbyrne*

48