Why be vague?

Kees van Deemter & Matt Green

University of Aberdeen
Scotland, UK

Herrenhausen, April 2013

Plan of the talk

- **Natural Language Generation (NLG)**
- **When do hearers benefit from vagueness?**
- **Experimental work**
  - E.Peters et al.
  - Mishra et al.
  - Green & van Deemter

Plan of the talk

- **Natural Language Generation (NLG)**
- **When do hearers benefit from vagueness?**
- **Experimental work**
  - E.Peters et al.
 - Mishra et al.
  - Green & van Deemter
- **Are we asking the wrong question?**

Natural Language Generation (NLG)

- **Goal: generate sentences/texts**

  For practical use, or
  to model human language production

- **Input:**

  Some **non-linguistic** information (database records, logic formulas, program code, …)

- **Example involving practical use:**

  FOG system for weather forecasting

  Operational since 1992

FoG: Input (as displayed on a map)
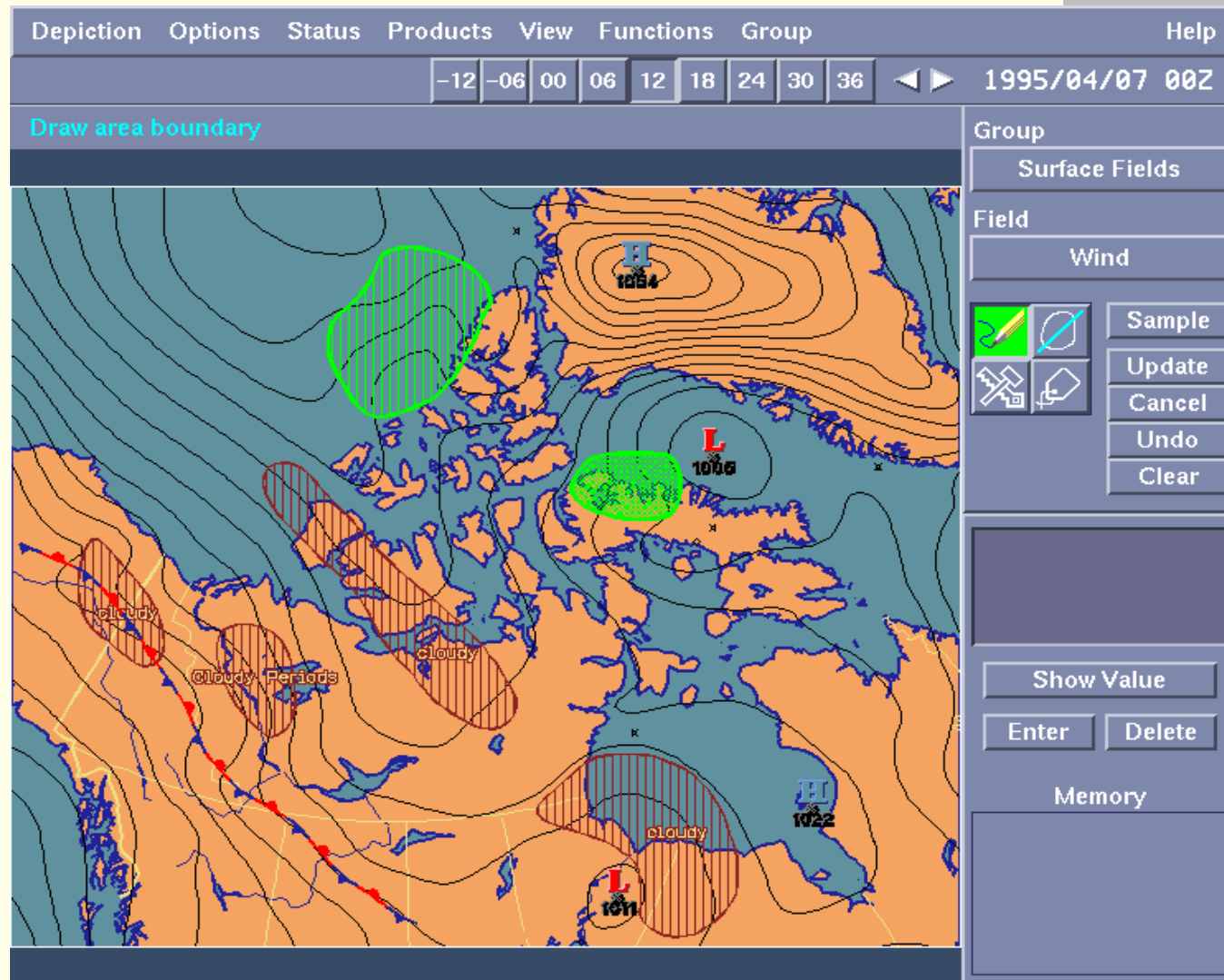
FoG system output (1992)


```
FPCN20  Status: CURRENT-NOT RELEASED                          Forecasts

FPCN20 CWEG 152300                                           -Marine--
MARINE FORECASTS FOR ARCTIC WATERS ISSUED BY THE ARCTIC WEATHER CENTRE   * ARWC **
OF ENVIRONMENT CANADA AT 05.00 PM MDT SATURDAY 15 APRIL 1995 FOR TONIGH   FPCN20
AND SUNDAY WITH AN OUTLOOK FOR MONDAY.                                    FPCN21
THE NEXT SCHEDULED FORECAST WILL BE ISSUED AT 05.00 AM MDT.               FPCN22/74
WINDS ARE IN KNOTS.                                                       FPCN23/75
FOG IMPLIES VISIBILITY LESS THAN 5/8 NM.                                  FPCN24/76
MIST IMPLIES VISIBILITY 5/8 TO 6 NM.                                      FPCN25/77
                                                                          UL 22/83
GREAT SLAVE LAKE.                                                        -Public--
WINDS LIGHT TONIGHT AND SUNDAY. SNOW ENDING NEAR MIDNIGHT. VISIBILITIES   FPCN15
NEAR 2 NM IN SNOW.
OUTLOOK FOR MONDAY... LIGHT WINDS.

GREAT BEAR LAKE.                                                   Set Element Priority ...
FREEZING SPRAY WARNING ISSUED.
WINDS EAST 20 TO 25 TONIGHT AND SUNDAY. FREE    SPRAY.             Set Active Areas ...
OUTLOOK FOR MONDAY... WINDS EASTERLY 20 TO 25.

MACKENZIE RIVER FROM MILE 0 TO MILE 100.                          Source
WINDS LIGHT TONIGHT AND SUNDAY. SNOW ENDING THIS EVE    SIBILITIES    Working Version
NEAR 2 NM IN SNOW.                                                    Official Release
OUTLOOK FOR MONDAY... LIGHT WINDS.                                   Forecast Rollup

MACKENZIE RIVER FROM MILE 100 TO MILE 300.                        Language
WINDS LIGHT STRENGTHENING TO SOUTHEAST 15 SUNDAY AFTERNOON. S
EARLY THIS EVENING. VISIBILITIES NEAR 2 NM IN SNOW.                   glish
OUTLOOK FOR MONDAY... WINDS SOUTHEASTERLY 15.

Generate    Update    Edit ...    Release
```

"Great Slave Lake: *Winds light tonight and Sunday, snow ending near midnight. Visibilities near 2NM in snow ...*"

NLG systems expressing quantities

Weather forecasting

<u>Input</u>: numbers (20 Knots, 11PM)

<u>Output</u>: *"Winds light tonight; snow ending near midnight"* (FoG system, Goldberg et al. 2000)

Medical decision support

<u>Input</u>: Time-series data on babies in IC

<u>Output</u>: *"about 0.3 litres, … very variable, …"*

(Babytalk system, Portet et al. 2012)

Vague expressions abound

"BREATHING – Today he managed 1½ hours off CPAP in about 0.3 litres nasal prong oxygen, and was put back onto CPAP after a desaturation with bradycardia. However, over the day his oxygen requirements generally have come down from 30% to 25%. Oxygen saturation is very variable. Usually the desaturations are down to the 60s or 70s; some are accompanied by bradycardia and mostly they resolve spontaneously, though a few times his saturation has dipped to the 50s with bradycardia and gentle stimulation was given. He has needed oral suction 3 or 4 times today, oral secretions are thick."

[BT-Nurse scenario 1]

Vagueness

- Vagueness as defined by logicians/philosophers/linguists:

  - Predicate is vague if it has **borderline cases**
  - (and borderline cases of borderline cases, causing sorites paradox)

- Is it ever helpful to be vague?
- Should practical NLG systems use vague language?

Why is language vague?

Barton Lipman: Why have we tolerated an apparent "worldwide several-thousand year efficiency loss"?

In A.Rubinstein, *"Economics and Language"* (2000)

Lipman's scenario

Airport scenario: *I describe Mr X to you, to pick up X from the airport. All I know is X's height; heights are uniformly distributed across people on [0,1].*
If you identify X right away, you get payoff 1;
if you don't, you get payoff -1

What description would work best?

- *State X's height "precisely"* $\Rightarrow$
  If each of us knows X's exact height then the probability of confusion is close to 0.

What description would work best?

- *State X's height "precisely"* $\Rightarrow$
If each of us knows X's exact height then the probability of confusion is close to 0.

What description would work best?

- *State X's height "precisely"* ⇒
  If each of us knows X's exact height then the probability of confusion is close to 0.

  Lipman: no boundary cases, hence not vague

  Theorem: under standard game-theory assumptions (Crawford/Sobel), vague communication can never be optimal

- When is vague communication more useful than crisp communication?
- **Strategic** vagueness. This is <u>not</u> about situations where the speaker has no choice (e.g., where no exact metric exists)
- Focus on **collaborative** situations

  Van Deemter (2009) "Utility and Language Generation: The Case of Vagueness". *Journal of Philosophical Logic* **38**/6.

Experimental work

1. Peters et al. (2009)

Herrenhausen, April 2013

Peters et al. (2009)

Hospital ratings based on numbers:

(1)   survival %
(2)   % of recommended treatment
(3)   patient satisfaction

"How attractive is this hospital to you?"

Peters et al. (2009)

- When labels ("fair", "good", "excellent") were added, a greater proportion of variance in evaluation judgments could be explained by the numeric factors

- Without labels,
  - the most important information (survival %) was not used at all
  - less numerate subjects were influenced by mood (*"I feel good/bad/happy/upset"*)

Herrenhausen, April 2013

Peters et al. (2009)

- This looks like a benefit from vague words ("fair", "good", "excellent")

Peters et al. (2009)

- This looks like a benefit from vague words ("fair", "good", "excellent")

- But …

  - The effect was caused by **evaluative** words
  - Nothing to do with borderline cases

    A vertical bar was used as a threshold

Peters et al. (2009): stimuli

Figure 1. Study 1: Hospital information is provided in an evaluative-categories format (top) or with numbers only (bottom).

2. Mishra et al. 2011

- How does feedback affect behaviour?
- Subjects wanted to loose weight

Group 1 were given precise feedback:
BMI = x

Group 2 were given a range:
x ≤ BMI ≤ y

Algorithm a says BMI=x
Algorithm b says BMI=y

2. Mishra et al. 2011

- Results: more weight loss in **Group 2**
- Mishra's explanation: using a range allows subjects to feel optimistic about their progress
  - Wishful thinking is common (≥ Marks 1951)
  - Feeling near one's target helps performance

- Interesting!

2. Mishra et al. 2011

… but:

Vagueness didn't play a role:

- Mishra's range had no borderline cases

- Explanation didn't involve borderline cases

- The real issue was **low granularity** (cf. Hobbs 1985)

3. Green & van Deemter

- Focus on referring expressions
- Compare readers' Response Times, e.g.

    Choose the square with four dots

    Choose the square with many dots

- Are RTs smaller for vague instructions?

"Choose the square with ... dots"
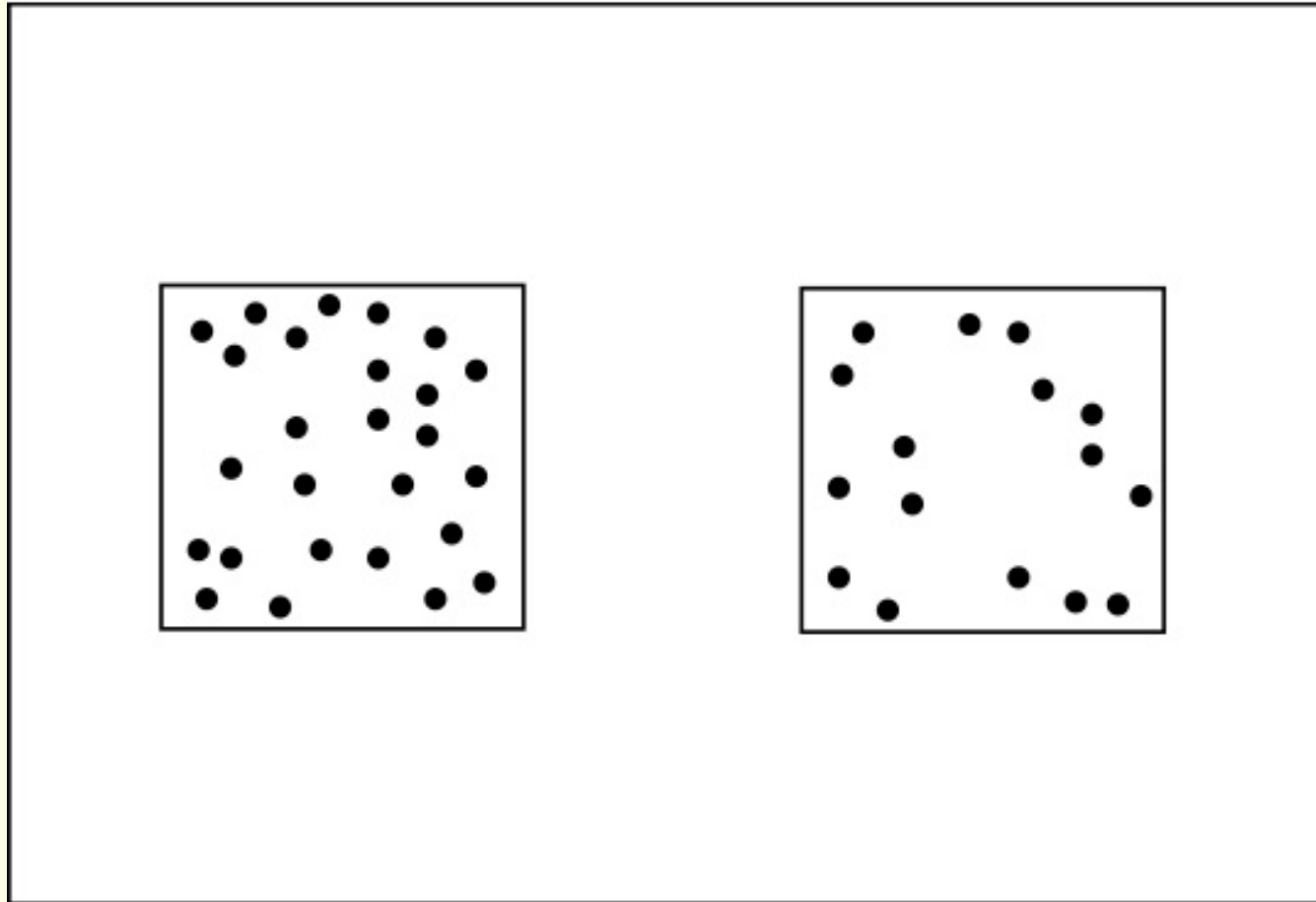
G&vD: Experiment 1

- No significant effect of vagueness

- **Subitizable** numbers followed the opposite pattern

- **Numbers below 5** play special role in visual perception, e.g.,

    Kaufman et al. (1949)

    Trick & Pylyshyn (1994)

G&vD: Experiment 1

Results for stimuli with a subitizable number of dots
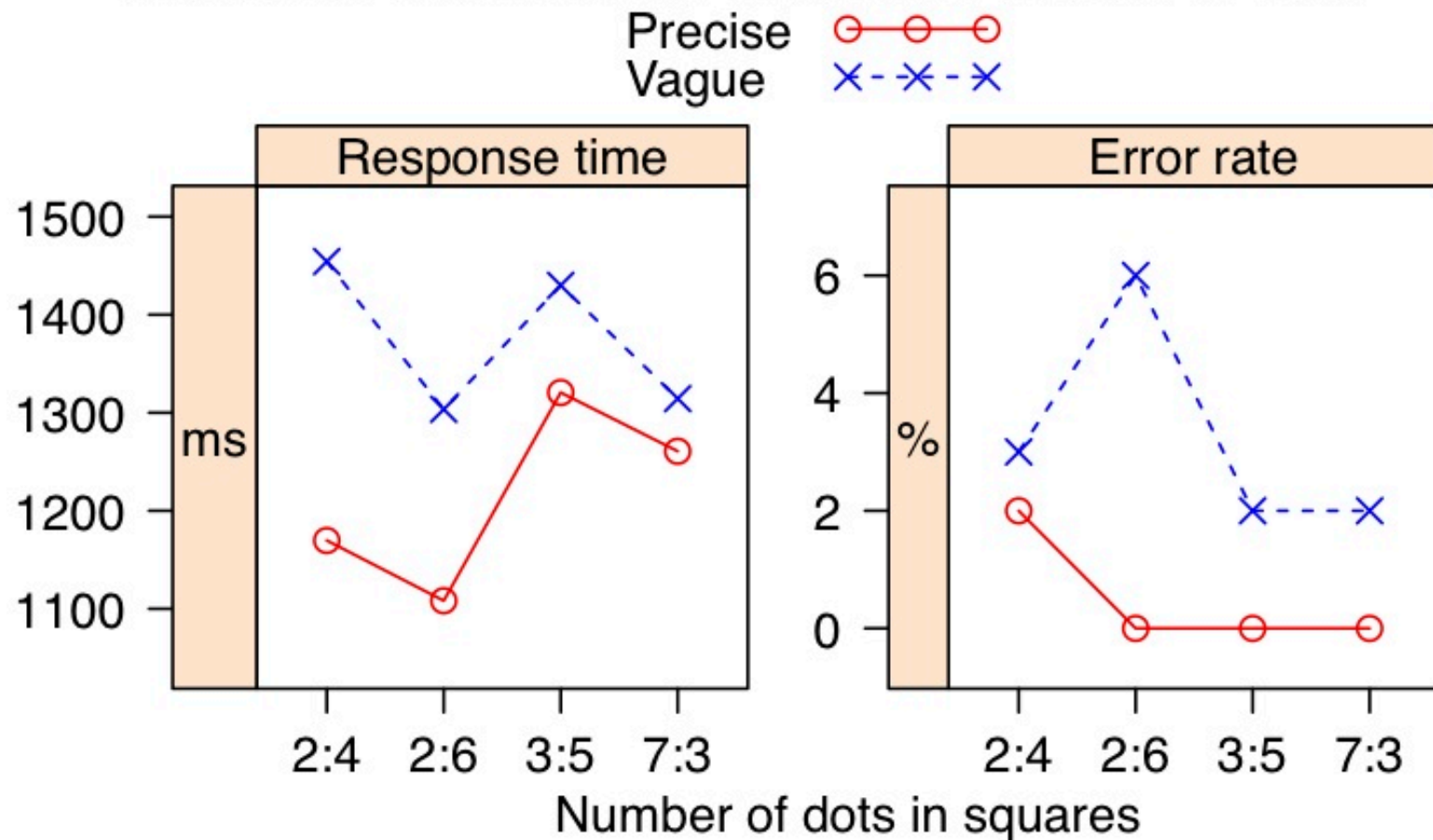
G&vD: Experiment 2

We varied

- the numbers of dots in the boxes: 5,10,15,20,(25),30,35,40,45

- The distance between the two numbers 5,10,15,20

All boxes were compared with

a box that contains **25** dots

G&vD: Experiment 2
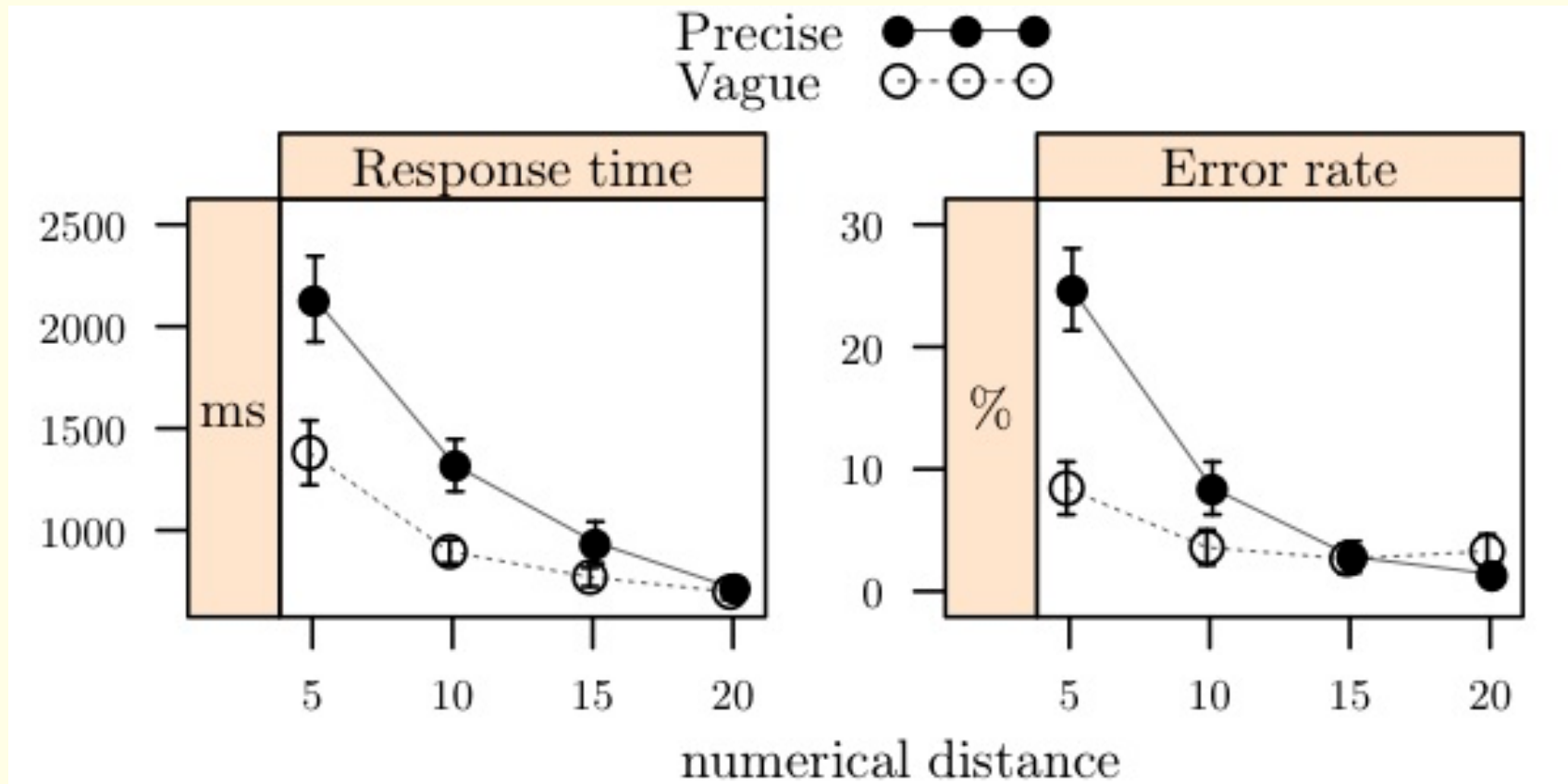
G&vD: Experiment 2

- Vagueness helps for larger numbers
- Diminishing advantage for vagueness as gap size grows large
  - Subjects are able to pick *"the square with **45** dots"* without counting

G&vD: Experiment 2

- Vagueness helps for larger numbers
- Diminishing advantage for vagueness as gap size grows large

- These were encouraging results, but …

Problems with Experiment 2

- Potential for vagueness not realised?
  - Two squares → no borderline case
  - Definite NPs (*"the square with .."*) identify the target uniquely
- Solution
  - use > 2 squares
  - use indefinite articles (*"a square with .."*)

G&vD: Experiment 3
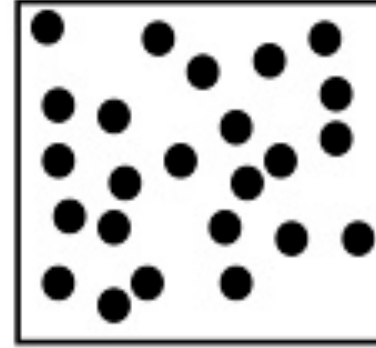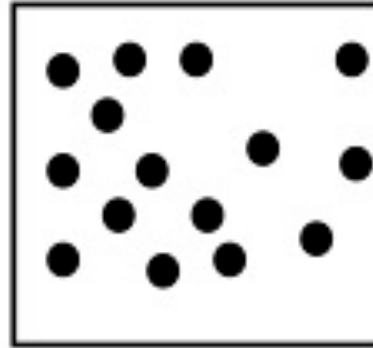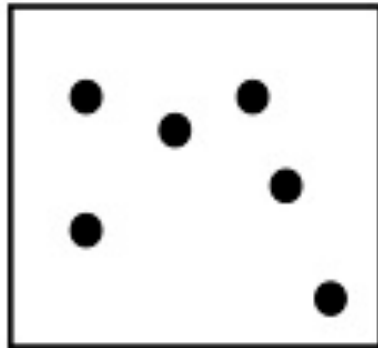
Other problems with Experiment 2

- Vagueness confounded with absence of numbers?

  Vague: *few, many*
  Precise: *5, 25*

- Solution: factorial instruction format (2 x 2)

	Precise	**Vague**
Numerical	…	…
Verbal	…	…

Example: The triple (16,25,34)

	Precise	**Vague**
Numerical	*16 dots*	*about 20 dots*
Verbal	*the fewest dots*	*few dots*

"a square with about 20 dots":

a clear case: 16 dots

borderline case: 25 dots

Example: the triple (16,25,34)

Vague & Numerical: *"about 20"* and *"about 30"*
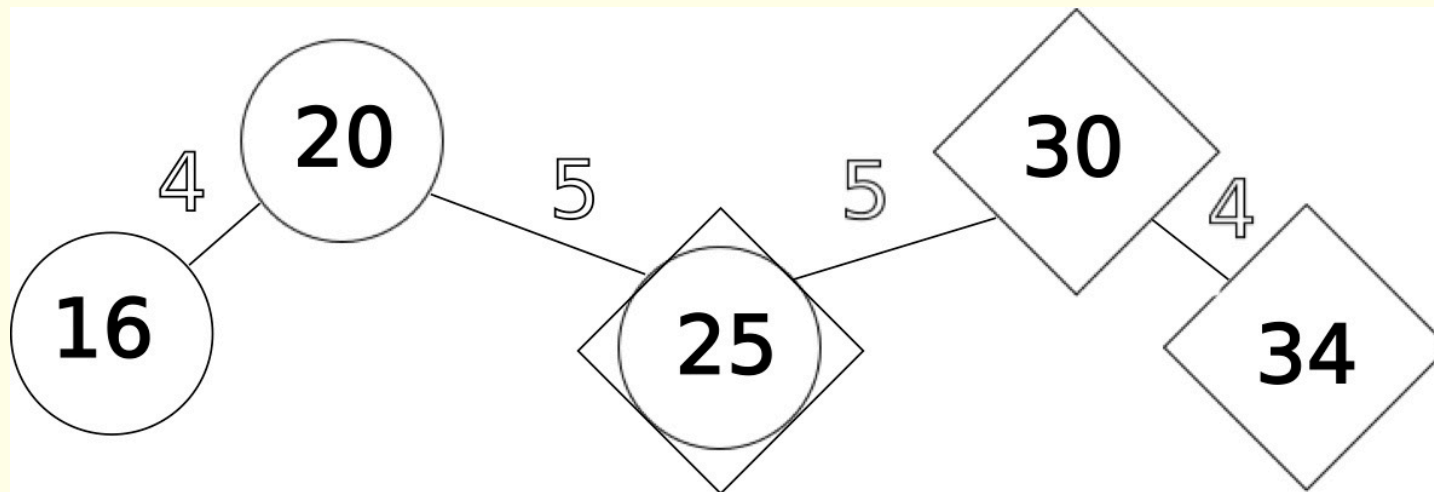
G&vD: Experiment 3 (1)
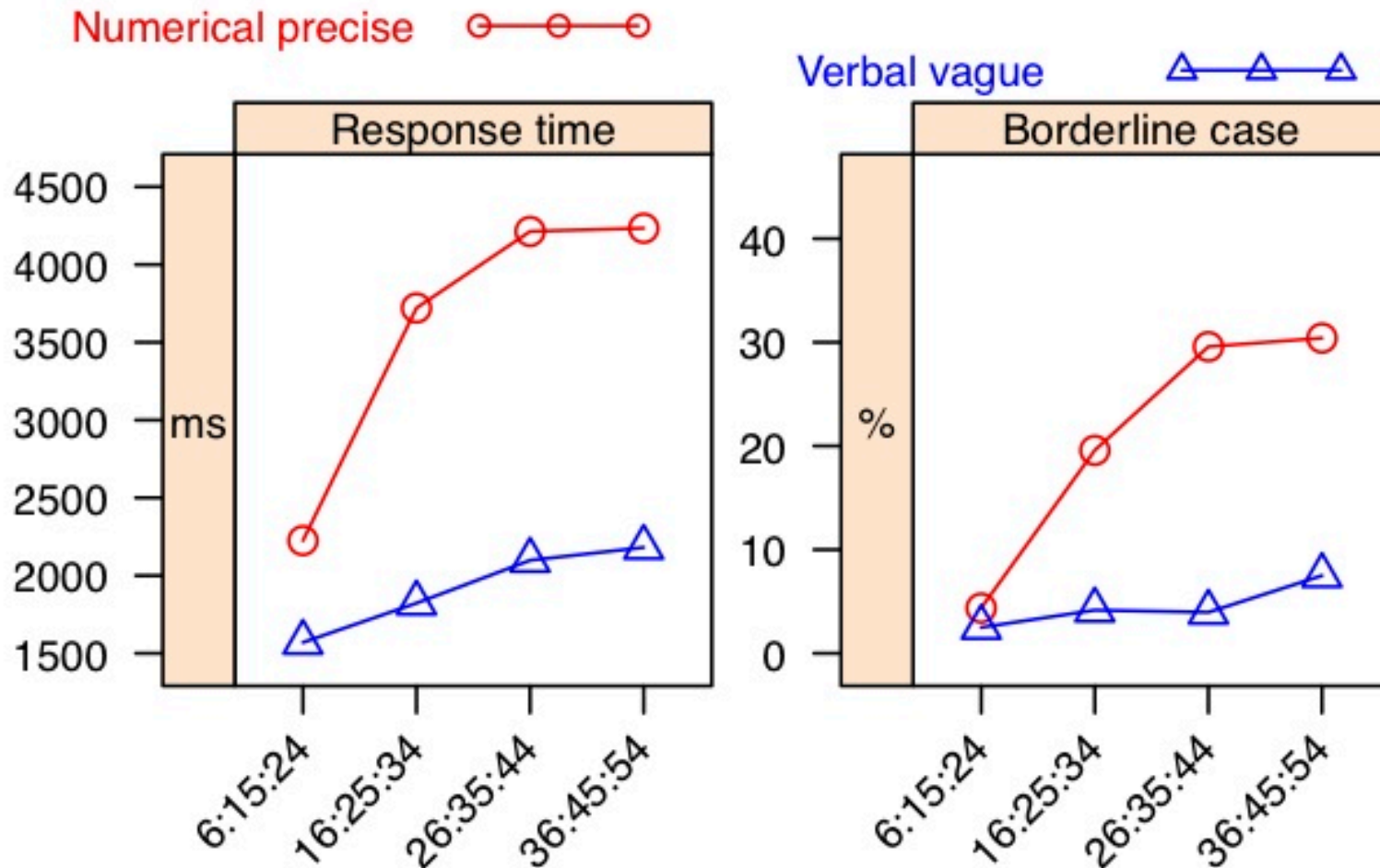
G&vD: Experiment 3 (3)
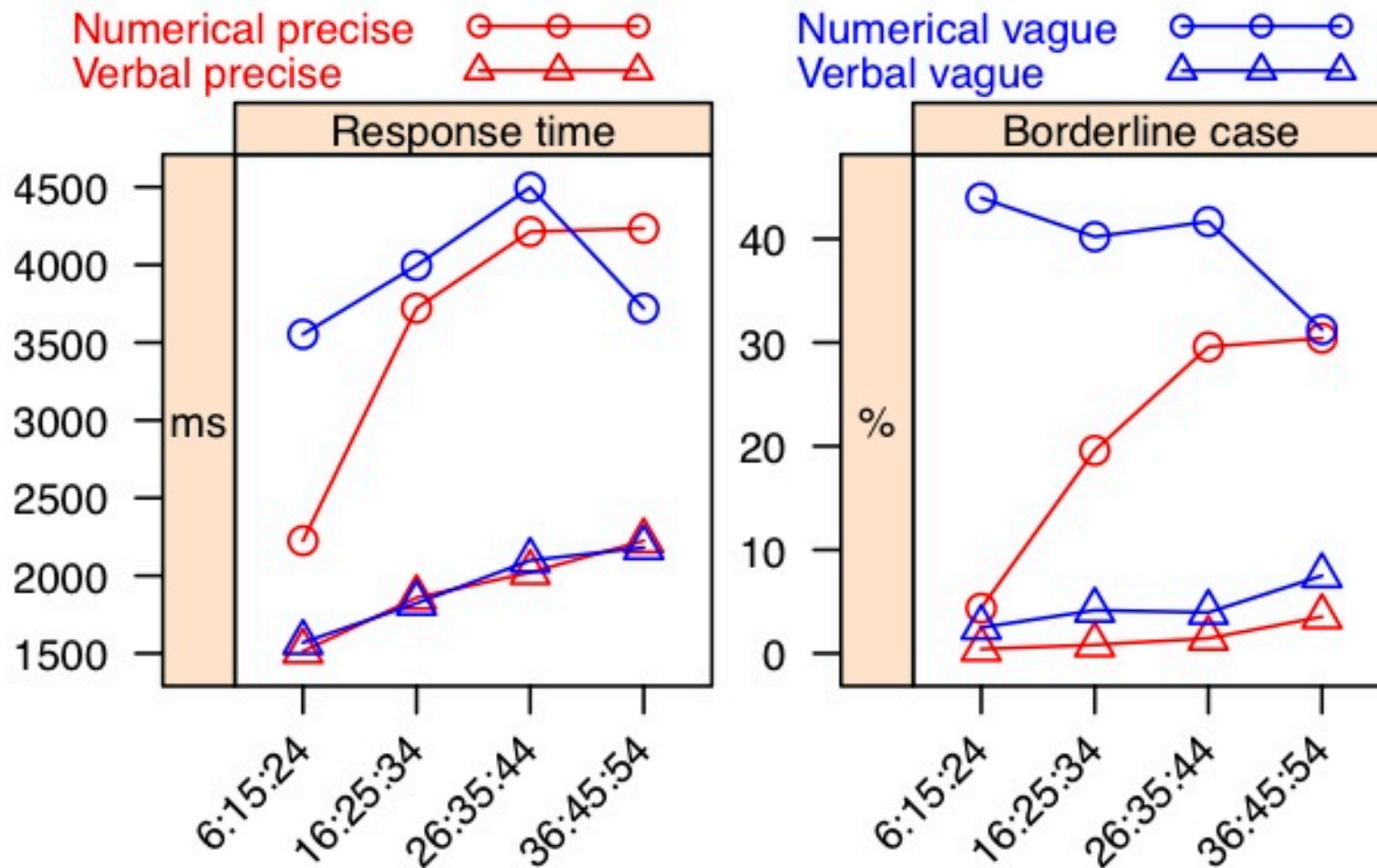
G&vD: Experiment 3

- **Vague** expressions were only marginally faster than Non-vague ones (p=0.73)

- **Numerical** expressions were much slower than non-numerical ones

- No interaction between vagueness and numericity

Herrenhausen, April 2013

What this suggests

- Perhaps the benefits of vague words are not about vagueness but about number avoidance

- Vague expressions are only better than crisp ones because they **tend** to

    - avoid numbers (Green & van Deemter)

    - express value judgments (Peters et al.)

    - have low granularity (Mishra et al.)

Alternative explanation

- But maybe it's not the presence of numbers *per se* .. either
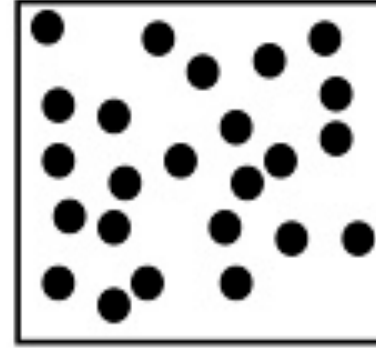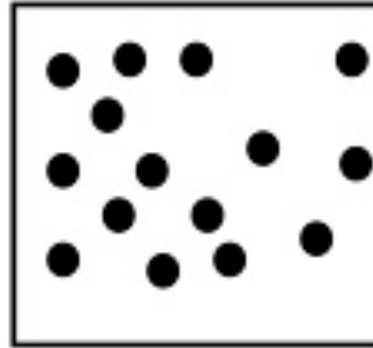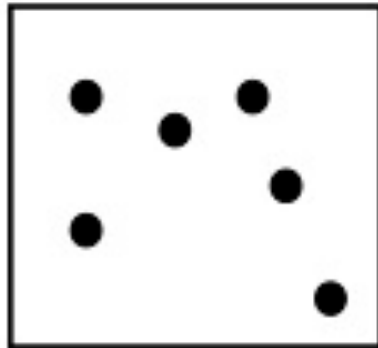
Example: The triple (16,25,34)

	Precise	**Vague**
Numerical	*16 dots*	*about 20 dots*
Verbal	*the fewest dots*	*few dots*

Both verbal items rely on a comparison task

Both numerical items rely on a matching task

G&vD: Experiment 3

New experiment **a** (focussing on **Numerical** expressions)

	Precise	**Vague**
Comp	*fewer than 30*	*far fewer than 30*
Match	*16*	*about 20*

- Even though the two **Comp** items contain a number, the task can be performed by finding the smallest number

Herrenhausen, April 2013

New experiment **b** (focussing on **Non-numerical** expressions)

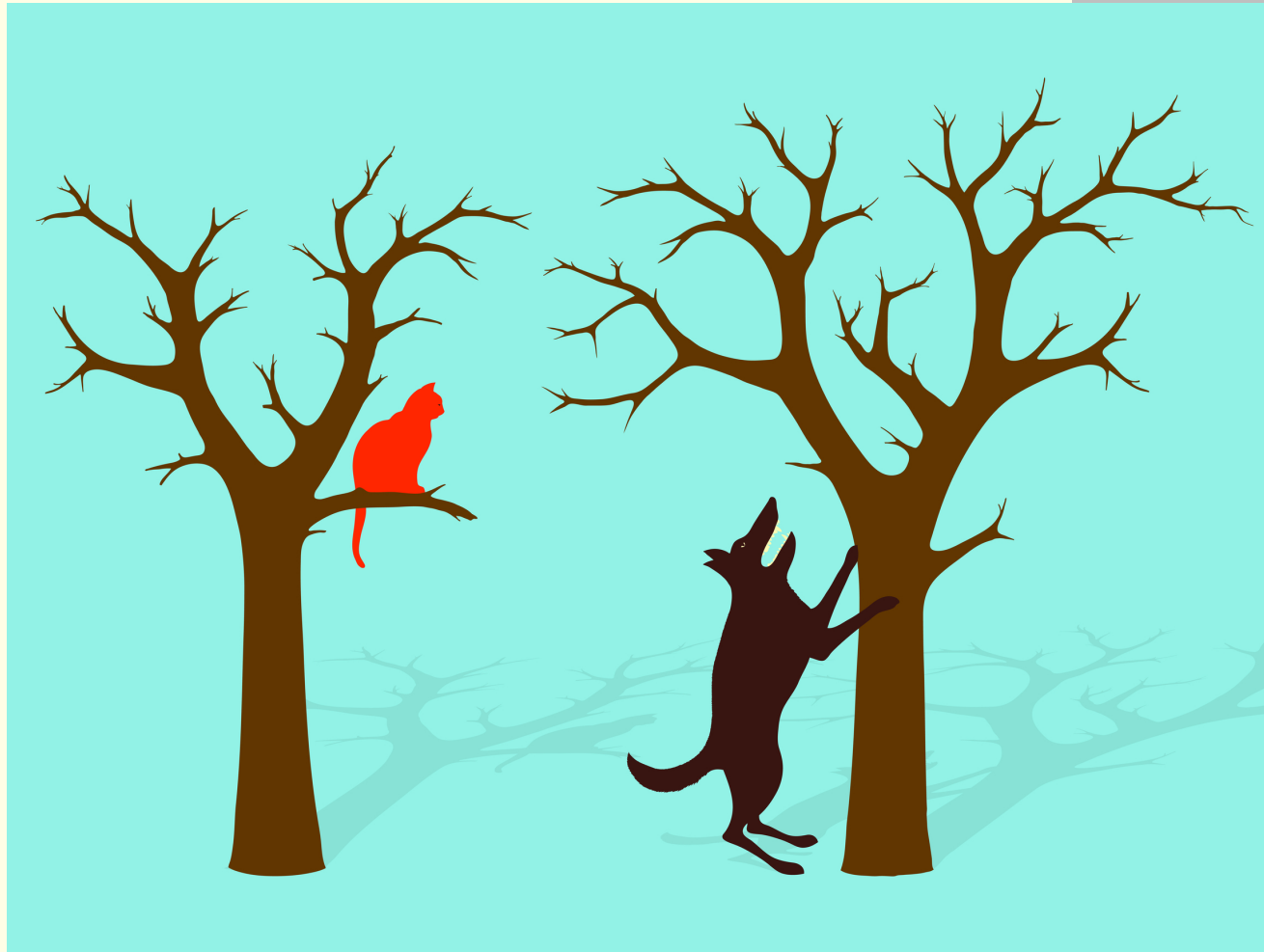|	**Precise**	**Vague**
Comp	*fewer than X*	*far fewer than X*
Match	*same number as X*	*approx same number as X*

- *X* has been shown a few seconds earlier
- Outcome of this experiment (**a** and **b**) not known yet
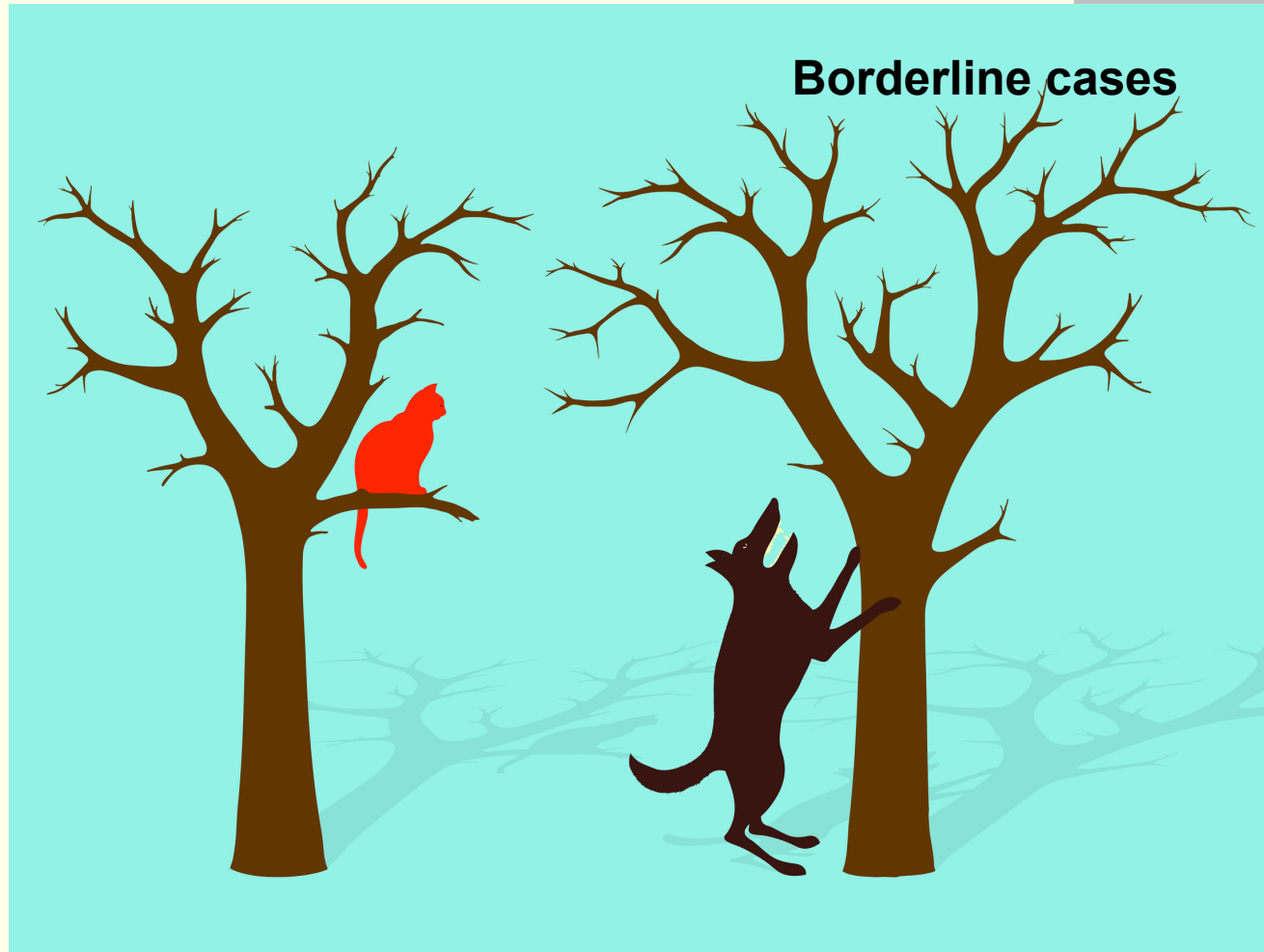
Where are we now?

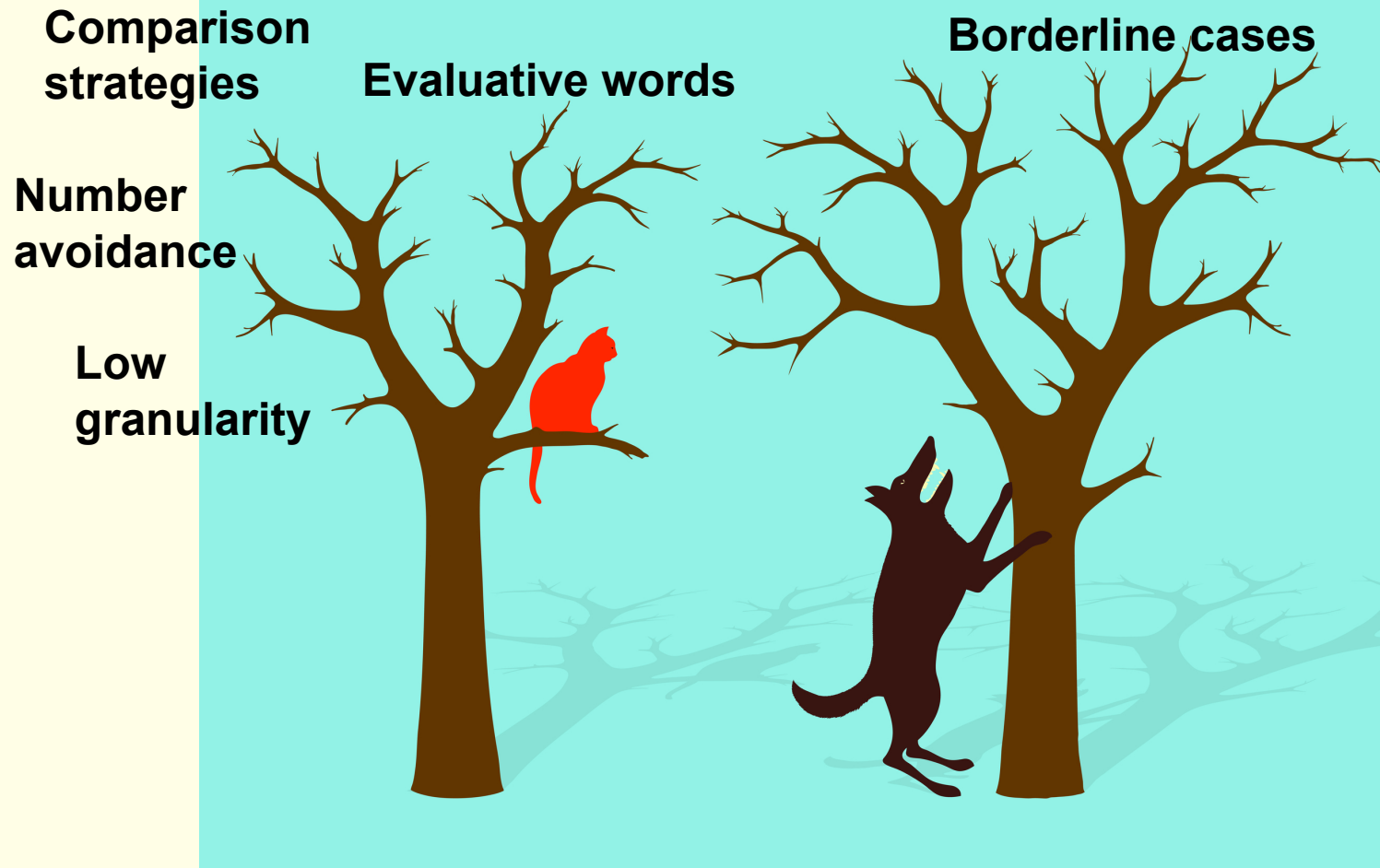Many of us are asking what's the use of vagueness.

Are we barking up the wrong tree?

Herrenhausen, April 2013

Are we barking up the wrong tree?

Herrenhausen, April 2013

Are we barking up the wrong tree?

Comparison strategies

Number avoidance

Low granularity

Evaluative words

Borderline cases

Herrenhausen, April 2013

Questions?

■ Green & Van Deemter (2011). Vagueness as Cost Reduction: an Empirical Test. In Proceedings of the workshop *Production of Referring Expressions*, CogSci 33.

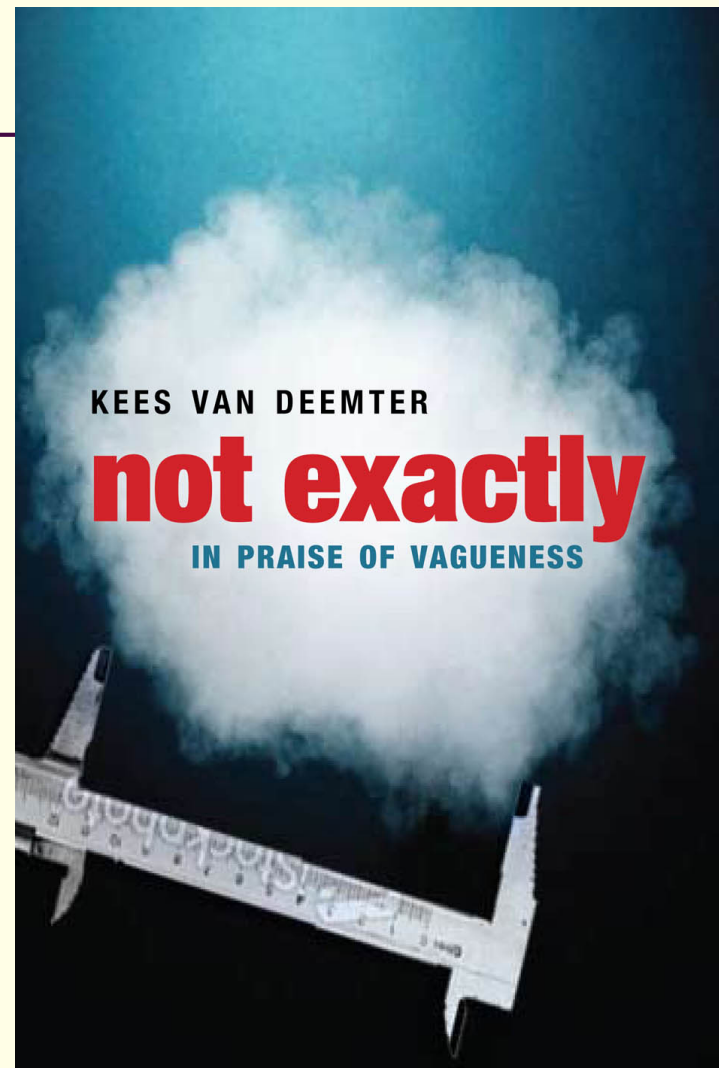**"Not Exactly: in Praise of Vagueness"**
Oxford University Press, Jan. 2010

<u>Part 1</u>: Vagueness in science and daily life

<u>Part 2</u>: Theories of vagueness

<u>Part 3</u>: Vagueness in Artificial Intelligence

www.csd.abdn.ac.uk/~kvdeemte/NotExactly

Herrenhausen, April 2013

1. Production/interpretation effort

- **Example**: during the planning of a trip:

  1. *The temperature is 23.75 C*
  2. *It's mild*

- (2) takes fewer syllables than (1)
- Precision of (1) adds little benefit
- E.g., Feasibility of an outing does not depend on whether it's 20C or 30C

2. Evaluation payoff

- <u>Example</u>: The doctor says
  1. *Your blood pressure is 153/92*
  2. *Your blood pressure is high*

- (2) offers less detail than (1)
- But (2) also offers **evaluation** of your condition (cf. Veltman 2000)
  - A link with actions (cut down on salt, etc.)
  - Crucial if metric is "difficult"

3. Comparing vs. measuring

Herrenhausen, April 2013

Comparing vs. measuring

- Example: One house of 11m height and one house of 12m height

  1. *the house that's 12m tall needs to be demolished*

  2. *the tall house needs to be demolished*

- Comparison is easier and more reliable than measurement ➔ prefer utterance 2

4. Future contingencies

- **Indecent Displays Control Act (1981) forbids display of *indecent matter***
    - "indecent" <u>at the time</u>

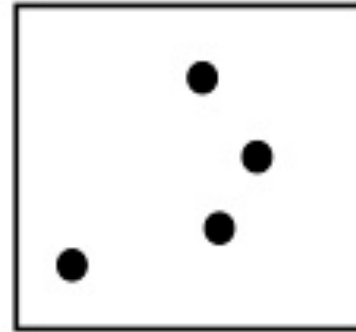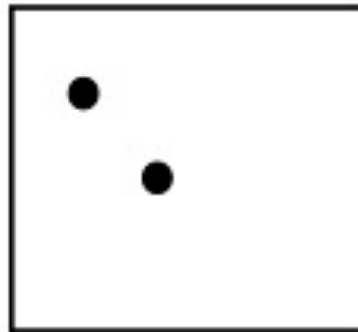 $\Rightarrow$ the law has been parameterised

    Waismann 1968, Hart 1994

- **Game-theoretical accounts**

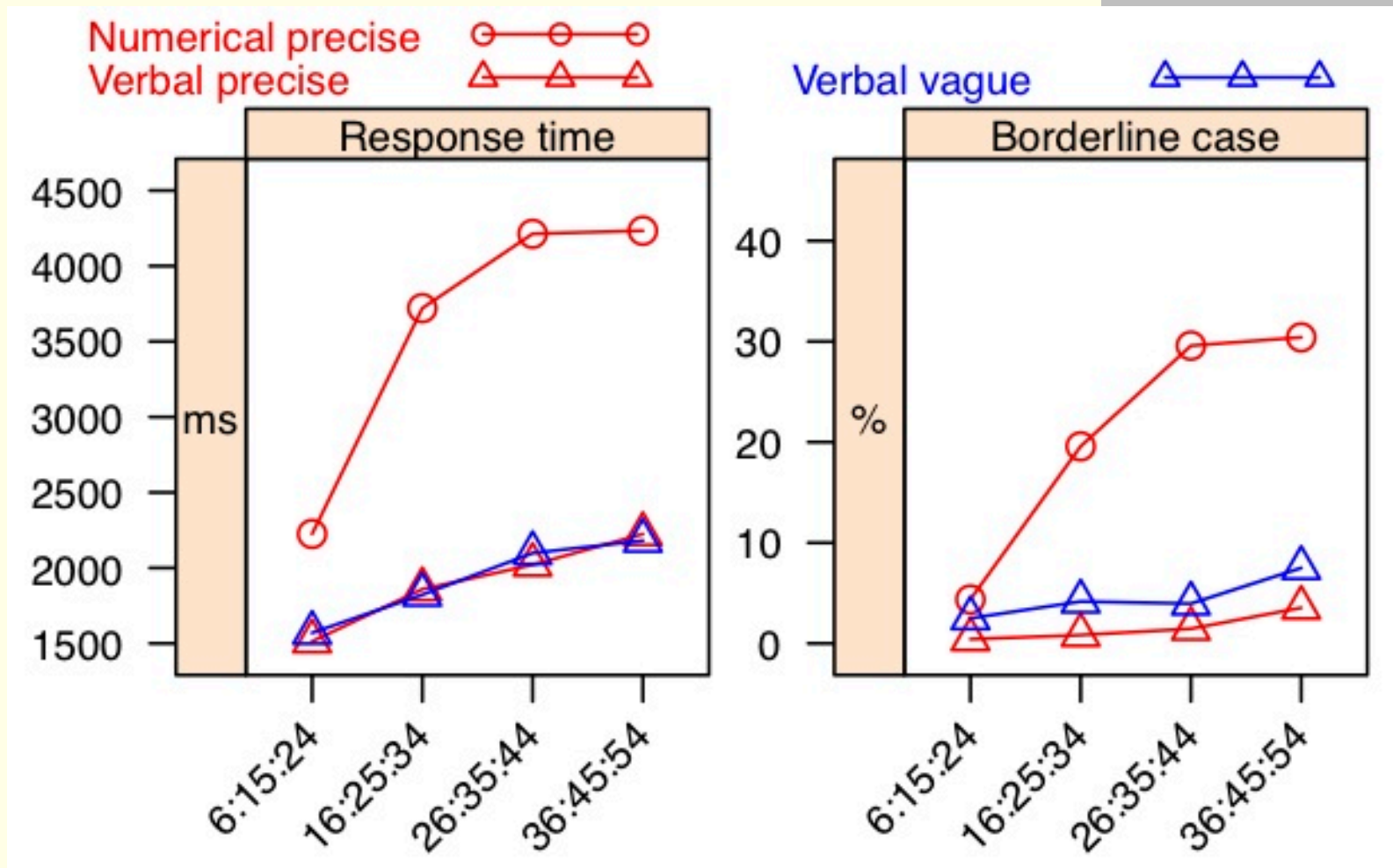   Aragones & Neeman 2000

G&vD: Experiment 1 (pilot)

Choose the square with four dots

G&vD: Experiment 3 (2)

G&vD: Experiment 3

number of dots in the display