Approximate and Precise Use of Language

Conference: Vague Language — Vague Law?
Albert-Ludwigs-Universität Freiburg
January 20 – 21, 2012

Manfred Krifka
krifka@rz.hu-berlin.de

Zentrum für Allgemeine Sprachwissenschaft, Berlin*
Humboldt-Universität zu Berlin

*Funded by BMBF, DFG, ESF

Overview

What I would like to show:

- Apparently precise expressions can be used in an approximate way.
- Apparently vague expressions have a precise core.
- Vagueness is perhaps not a feature of language structure (semantics), but it is definitely a feature of language use (pragmatics).
- By allowing for vagueness in language use, language increases its usefulness in communication.
- Allowing for vagueness in language structure as well would probably have detrimental effects.

What I cannot discuss for lack of expertise:

- The role of vague language in legal discourse.

But I will refer to aspects of legal discourse where I think it is appropriate.

Topic I:
Approximate interpretation of precise expression

- Prototypical precise expressions: Number words.
  There were fifty-three people at the vagueness conference.
  true iff #{x | x is a person at the vagueness conference} = 53,
  false if = 52, or = 54
- But number words can be interpreted approximately:
  There were fifty people at the vagueness conference.
  not judged false in case there were 46 or 53 persons.
- Other examples:
  The line is forty centimeters long.
  The line is forty point zero centimeters long.
  The train had a delay of fifteen minutes.
  The train had a delay of seventeen minutes.
  We had to wait for an hour and a half.
  We had to wait for ninety minutes.
  The street turns ninety degrees right.
  The street turns one hundred degrees right.

Questions:

- Are number words and measure expressions vague after all?
- What influences the precision of interpretation of number words?

Approximate interpretation of precise expressions: Some real-life cases.

Protest sign in Great Britain against the introduction of the metric system

Street sign in Kloten, Switzerland

— *How old are these bones, Paps?*
— *Believe it or not, 65,000,0017 years!*
— *How come?*
— *Simple math., Son. They told me they were 65,000,000 million years when I visited the museum the first time, and that's now 17 years ago.*

Are number words vague?

A not very attractive suggestion:

- Number words denote ranges:
  ⟦*fifty*⟧ = [45...55], ⟦*fifty-three*⟧ = [53]

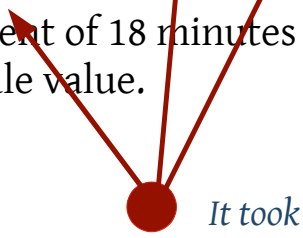
An alternative proposal (Krifka 2008):

- Number words are precise, but...
- they come with alternatives that form scales, and...
- these sets of alternatives are ordered according to their granularity, where...
- the size of an entity is identified with the closest number w.r.t a scale.

Example: three minute scales

```
a. -10-11-12-13-14-15-16-17-18-19-20-21-22-23-24-25-26-27-28-29-30-
b. -10-------------15-------------20-------------25-------------30-
c. -10-------------15-------------------------------------------30-
```

Measurement of an event of 18 minutes with respect to these scales:
Mapping to closest scale value.



It took eighteen / twenty / fifteen minutes

Principles of scale coarseness

- The nature of the counting system:
  - mulitples of powers of ten in a decimal language:
    ten, twenty, thirty, ...
    one hundred, two hundred, three hundred, ...
  - multipes of 12 in a duodecimal scale:
    one dozen, two dozen, three dozen
- halving existing scales:
  - halving the decimal scale
    five, ten, fifteen, twenty, twenty-five
  - halving other units
    quarter, half-dollar, dollar
    one hour, thirty minutes, fifteen minutes
    360 degrees, 180 degrees, 90 degrees
- Logarithmic scales
  kilobytes, megabytes, gigabytes, terabytes
  hamlet, village, town, city, (metropolis)
- Mixture of linear / logarithmic scales (cf. Hobbs 2000):
  five minutes, ten minutes, fifteen minutes, half an hour, an hour, an hour and a half

The pragmatics of coarseness: A preference for approximate language use?

One type of explanation:

- Coarse interpretations are preferred over precise ones.
- On hearing *fifteen minutes*,
  assume the coarsest scale at which *fifteen minutes* is represented
  (other considerations might apply,
  e.g. *we talked for fifteen minutes* vs. *the train will arrive in fifteen minutes*)

But: Why should there be a preference for coarse representations?

Possible reasons:

- Simplified cognitive representation
- Simplified inference processes
- Saving face – being approximate allows to be truthful;
  cf. Pierre Duhem, balance between precision and certainty
  cf. Elinor Ochs Keenan, non-informativeness in rural cultures

Yet: Do we really have to assume a pragmatic tendency to be approximate?

Selection of coarseness by strategic communication

We don't have to assume a pragmatic preference for coarseness!

Assume:

- Approximate and precise interpretations are equally likely,
  i.e. coarse and fine-grained scales are equally likely,
  (certain contexts might favor fine-grained scales – e.g., taxes)
- Principle of strategic communication (Prashant Parik):
  - Assume the most likely interpretation,
    given the choice of expression and a-priori likelyhood of message.
    A man is robbed in New York City every ten minutes.
    Likely reading: Every ten minutes there is a man that is robbed.
    Unlikely reading: There is a man that is robbed every ten minutes.

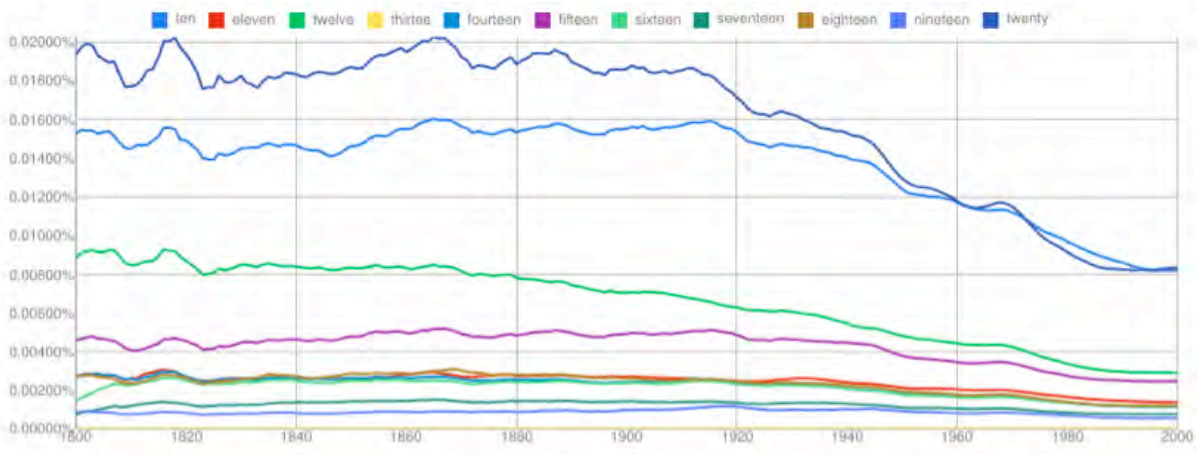
Selection of coarse scales by strategic communication

Derivation of preference for coarse scales:

- Assume: A-priori probability of durations of an event are roughly equal:
  p(13min) = r, p(14min) = r, p(15min) = r, p(16min) = r, etc.
- Possible scales where *fifteen minutes* denotes a value:
  a. *..., thirteen, fourteen, fifteen, sixteen, seventeen, ...*
  b. *five, ten, fifteen, twenty, twenty-five, ...*
  c. *five, ten, fifteen, thirty, fourty-five, ...*
- Speaker said: *fifteen minutes*
  - Assuming scale (a): Value 15min can be reported
    a-priori probability: p(15min) = r
  - Assuming scale (b): Value [13min...17min] can be reported
    a priori probability: p(13min...17min) = 5r (in any case, > p(15min)
  - Assuming scale (c): Value [10min...20min] can be reported
    a priori probability: p(10min...20min) = 10r (in any case, > p(13min...17min)
- Choosing the interpretation with the highest a-priority probability of event:
  - Select scale (c) over (b).
  - Select scale (b) over (a).

Reflection of coarseness in language: Frequency

Frequency of number word use:
number words *ten* to *twenty*, Google ngrams
(about 4% of all published English language books)



twenty > *ten* > *twelve* > *fifteen* > *eleven* > *thirteen, fourteen, sixteen* > *seventeen* > *nineteen*

Reflection of coarseness: Reduction of frequent expressions

Expressions that are frequently used tend to be shorter (Zipf's law), historical effect: shortening

- English: *fifteen (*fiveteen), fifty (*fivety)*:
  loss of diphthong, shortening
  OEng fi:f as word vs. fif- as prefix; vowel shift only affected i: (> ai),
  this did not affect *nine* in *nineteen*
- Phonological simplification of 'fifteen', 'fifty' in colloquial German:
  fuffzehn, fuffzig vs. regular *fünfzehn, fünfzig*:
  unrounding ü > u, loss of n, shortening (3 to 2 morae)
- Phonological simplification of 'half' in German
  anderthalb 'one and a half', lit. 'the second half' vs. regular *eineinhalb*
- Morphological simplification of 'fifty' in Danish:
  halvtreds vs. older *halvtredsind-s-tyve*

Influence of number system on language use:

The number system of a language might influence the approximate numbers that speakers use

- Norwegian (decimal)
- Danish (remnants of vigesimal, based on 20)
- Basque (fully vigesimal, based on 20)

Evidence: Google searches on domains for Norway, Denmark, Spain.

Num-ber	Norwegian		Danish		Basque	
	Numeral	#	Numeral	#	Numeral	#
20	*tjue*	61300	*tyve*	121000		
30	*tretti*	43700	*tredive*	25400	*hogei ta hamar*	892
40	*førti*	**39200**	*fyrre*	**26800**	*berrogei*	**85000**
50	*femti*	**81200**	*halvtreds*	**15500**	*berrogei ta hamar*	**213**
60	*seksti*	**19400**	*tres*	**36400**	*hirurogei*	**34000**
70	*sytti*	10200	*halvfjerds*	581	*hirurogei ta hamar*	69
80	*åtti*	13100	*firs*	3740	*larogei*	9000
90	*nitti*	13500	*halvfems*	540	*larogei ta hamar*	7

Coarseness level as a contextual parameter

The coarseness level stays the same in a text,
unless indicated otherwise:

At the vagueness conference, there were forty-seven participants,
and at the third-person-perspective conference, there were fifty.

- Precis interpretation of *fifty*.
  There were fifty participants at the third-person-perspective conference,
  and forty-seven at the vagueness conference.
- Garden-path-effect: *fifty* is interpreted as precise in hindsight.
  There were about seventy participants at the vagueness conference,
  and at the third-person-perspective conference, exactly fifty.

Possible consequences for legal language

Advise to the law-maker:
To improve strict compliance: Use non-round numbers!

Street sign in Slovakia

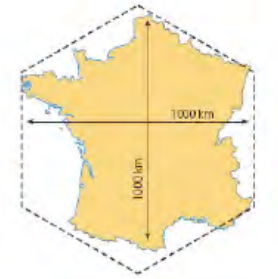
Where we stand, and where we go:

We have seen:

- Precise expressions (number words) can be used in approximate ways.

Other examples can be given, e.g.

France is hexagonal.



The head of this screw is hexagonal.

We will see:

- Apparently vague expressions can have a precise underlying meaning, vagueness derives from language use.

Vague Expressions: Gradable adjectives

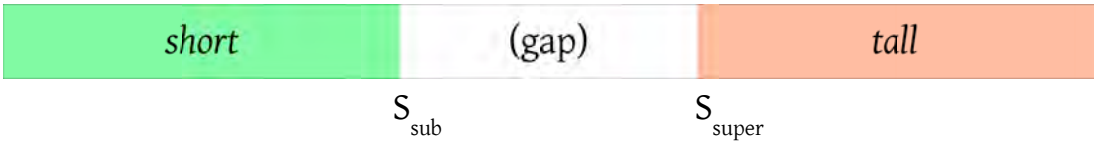
The prototypical case of vague expressions: Gradable adjectives

- Gradable adjectives allow for comparatives:
  positive: *tall* comparative: *taller*, superlative: *the tallest*, excessive: *too tall*
- They often come in pairs of antonyms:
  tall, short
  heavy, light
  dark, light
  happy, unhappy
- In certain dimensions, the lexical domains can be even richer:
  cold, cool, (tepid,) warm, hot

Semantics of gradable adjectives: Standard assumptions

Standard assumptions about gradable adjectives, antonym pairs:

- Their interpretation depend on a comparison class, often implicit.
  John is a tall ten-year-old.John is tall for a ten-year-old.
- Application of positive relative to a threshold value w.r.t. comparison class:
  HEIGHT(john) > S(HEIGHT(ten-year-olds))
- The threshold values for antonym pairs leave a gap (e.g., Klein 1980):
  John is neither tall nor small for a ten-year-old.
  - *John is tall for a ten-year-old*: HEIGHT(john) > S_{super}(HEIGHT(ten-year-olds))
  - *John is small for a ten-year-old*: HEIGHT(john) > S_{sub} (HEIGHT(ten-year-olds))



Problems of standard assumption:

- What precisely are the values of S_{sub} and S_{super} ?
- There isn't really a sharp boundary between *tall* and *not tall*.

Possible reactions towards problems of positive forms of gradable adjectives

- We give up interpretation with respect to two truth values {0, 1} and allow for interpretation with respect to truth-value range [0, 1] (e.g., fuzzy logic).
- We stick to precise semantic representation (truth values {0, 1}), and explain the setting of threshold patterns as an issue of language use.

Context-sensitive use of positive adjectives:

Kyburg & Morreau (2000):

A: *Which pig do you want?*
B: *I'll take the big one.*

A: *Which pig do you want?*
B: *I'll take the small one.*

Notice:
Even a fuzzy account of gradable adjectives would have to take context into account.

Assume:

- positive adjectives are interpreted precisely with respect to a parameter S:
  ⟦*tall*⟧ = {x | HEIGHT(x) > S(HEIGHT(comparison class))
- The context parameter can be set to maximize the usability of language.

A case study: Positives, negated positives, antonyms, and negated antonyms

happy	*not happy*
likely	*not unlikely*
intelligent	*not intelligent*
common	*not common*
successful	*not successful*
frequent	*not frequent*
many	*not many*

unhappy	*not unhappy*
unlikely	*not unlikely*
unintelligent	*not unintelligent*
uncommon	*not uncommon*
unsuccessful	*not unsuccessful*
infrequent	*not infrequent*
few	*not few*

negated antonyms,
"double negatives"

Cf. Horn 2002, "The logic of double negation"

Why are there double negatives?

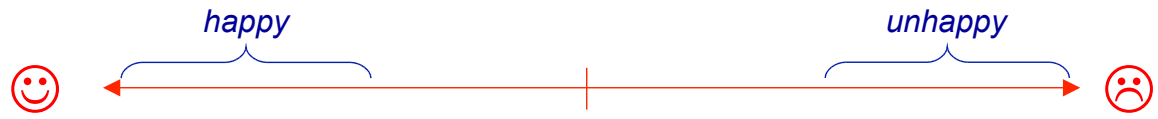
- Weakened meanings:
  - Jespersen (1924): The two negatives [...] do not exactly cancel one another, so that the result [*not uncommon, not infrequent*] is not identical with the simple *common, frequent*; the longer expression is always weaker: "this is not unkwown to me" or "I am not ignorant of this" means: 'I am to some extent aware of it', etc.
  - Fowler (1927): [Double negatives are] congenial to the English temperament [...] it is pleasant to believe that it owes its success with us to a stubborn national dislike of putting things too strongly.
- Pomposity:
  - George Orwell (1946): Banal statements are given an appearance of profundity by means of the *not un-* formation.
- Strengthened meanings:
  - Erasmus (1517, *Colloquia*): You shouldn't be left uninformed [!] that we use this sort of diction in two ways: for the sake of modesty, especially if we're talking of ourselves, and the sake of amplifying. For we say correctly and gracefully "not ungrateful" for "very grateful", "not vulgarly" for "singularly".
- Examples:

  For the first time in a long time I woke up feeling rested this morning. I felt rested and content and was not unhappy to find that my kids were already awake.

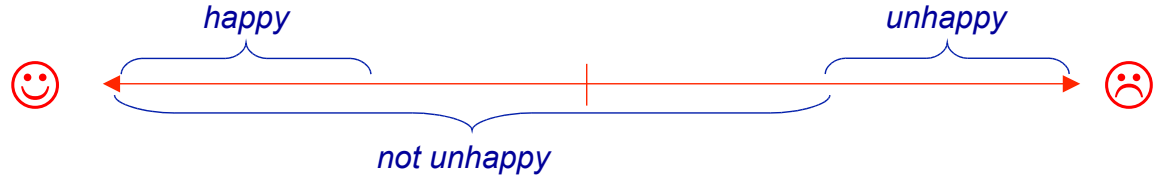
  For the most part he is well informed and not unintelligent.

Previous attempts to weakened positive meaning of negated antonyms: Horn (1991, 93), Blutner (2000)

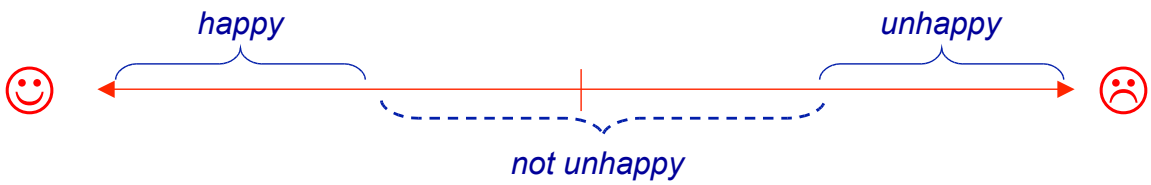
happy and *unhappy* are contraries, with standars S_{sub} and S_{super}.



Negation by *not* forms complements:



not unhappy is restricted to cases in which *happy* does not apply because it is more complex than *happy* (pragmatic competition)



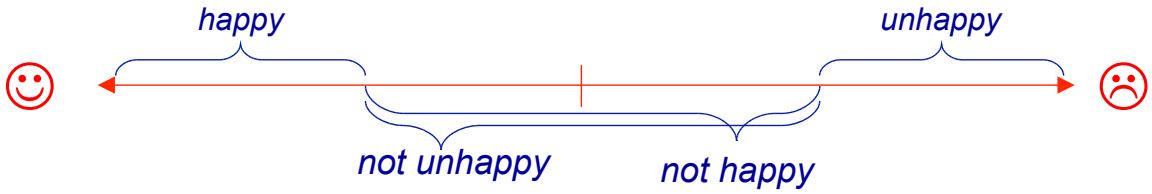
Two problems with this account

First problem: not *unhappy* ≠ *indifferent*:

cf. Judith M. Bardwick Ph.D., *The Plateauing Trap* (1986):

> "When we are plateaued, we are not so much actively unhappy as we are just not happy. We could continue to live as we are, because it's not awful. But it is also not joyous."

Second problem: *not happy* is interpreted differently from *not unhappy*, but predicted interpretation is the same.

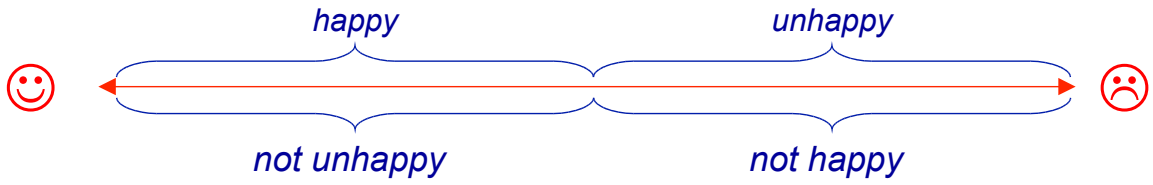


An alternative theory of negated antonyms (Krifka 2008):

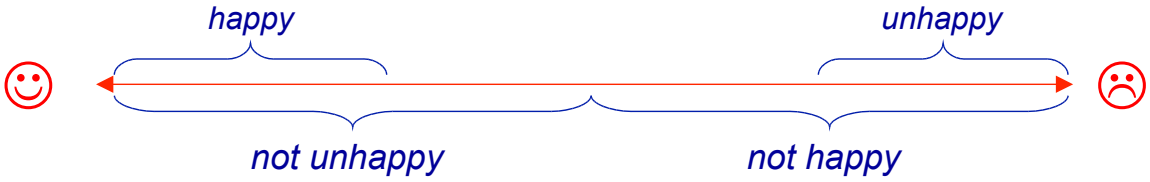
- Antonyms like *happy, unhappy* are **contradictories**, not contraries; they are exhaustive, i.e. there is **no gap** between their extensions.
- But they are preferably interpreted to refer to more extreme states of happiness and unhappiness, respectively.
  - The reason for this is pragmatic: Strenghthening of interpretation.
  - General type of pragmatic phenomenon:
    I-Implicature of simple expressions,
    M-Implicature of marked expressions. (cf. Horn, Levinson 2000)
  - Example: *She smiled.*
    Interpreted as: She performed a regular smile.
    She raised the corners of her mouth.
    Interpreted as: She performed an artificial, non-normal smile.
  - In the case at hand: Extreme setting of parameter S in order to minimize misunderstanding due to possible different setting of parameters; parameter setting to be on the safe side;
    cf. epistemic theory of vagueness (Williamson 1994).
- Consequently, the competing more complex expressions *not unhappy* and *not happy* are restricted to the remaining domain.

Illustration of negated antonyms
A Competition Theory for Happiness and Unhappiness

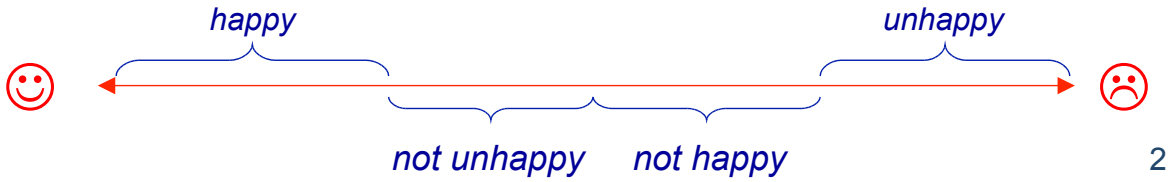
Initial situation: Antonym pairs and their negations.



Restriction of simpler expressions to clear cases.



Restriction of complex expressions to less clear cases.



How does parameter setting work?

Parameter setting maximizes communicative utility of language.
Some rules:

- Set the parameter in such a way
  that the addressee gets the intended meaning.
  (cf. Kyburg & Morreau, *the big pig / the small pig*).
- Assume a parameter setting that certainly corresponds
  to the parameter setting of your addressee --
  that is, better choose a more extreme setting of the parameter.
- Take the range of available expressions into account (granularity);
  different settings if expressions *tall, not tall, small, not small* are used.
- Choose a setting of the parameter that minimizes borderline cases.

Miminization of borderline cases

A common experience with grading exams:

```
Students                x       x       x
                        x  x    x       x   x           x
                     x  x  x    x       x   x   x       x   x
          x    x  x  x  x  x  x    x  x   x   x       x   x
Points    0–1–2–3–4–5–6–7–8–9–10–11–12–13–14–15
```

How would one assign this to grades, e.g. A – B – C – D – F?

Question:

- Do similar principles play a role in the setting of parameters of positive adjectives?

Experimental research on paramter setting

VAAG project: Vagueness, Approximation, Granularity (ESF, DFG)

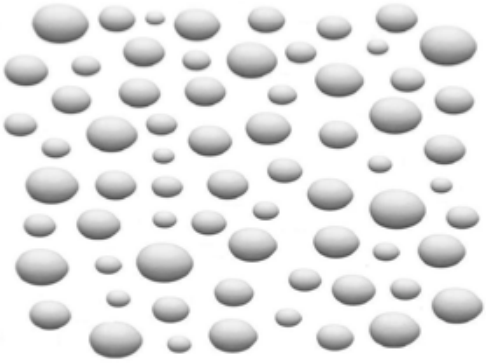
ZAS Berlin: Stephanie Solt, Uli Sauerland, Nicole Gotzner, Manfred Krifka

Work by Solt and Gotzner:

- How do people set parameters with gradable adjectives?
  - According to ranking (e.g., the one-third highest-ranked individuals), as proposed by Bale?
  - According to mean value, as proposed by Cresswell?
  - According to an (arbitrary) distance from the median value, as proposed by Solt?
- Does the distribution of the values of the objects have an influence on the parameter setting?
- Do previous attributions of properties have an influence on the parameter setting?

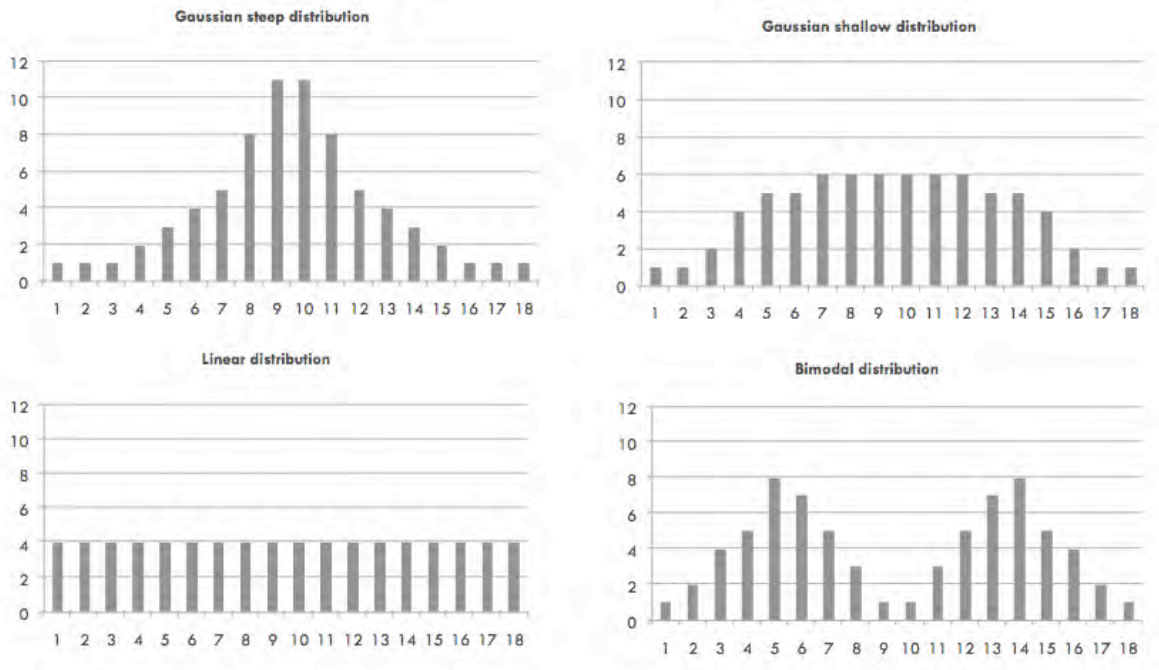
Experimental research on parameter setting

Experiments:

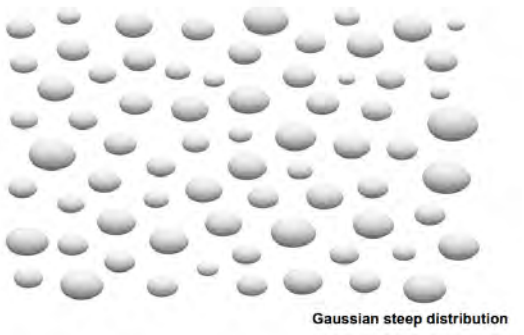


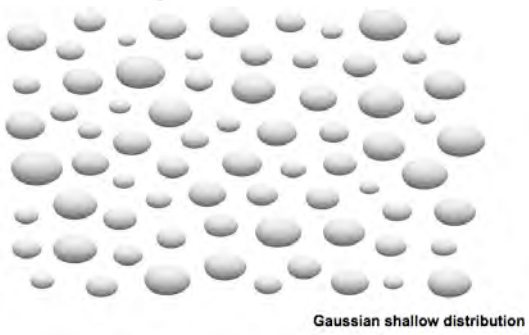
- Interpretation of adjectives w.r.t. an array of pictures representing the comparison class.
- For example: *Tick off the large eggs, and tick off the small eggs!*
- Varying the distribution, checking influence on judgements.
- Relevance of normal distribution for exemplars in species.
- Test with 77 native German speakers, 72 eggs, 18 sizes, four distributions: steep Gaussian, narrow Gaussian, linear, bimodal; task: mark the big eggs and mark the small eggs.

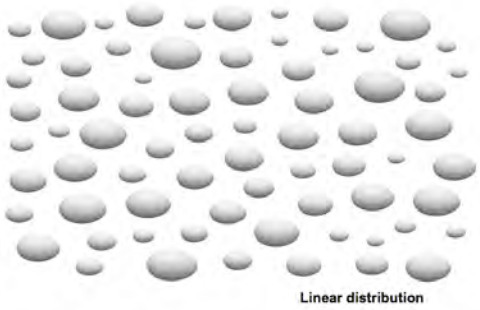
Experimental research on parameter setting

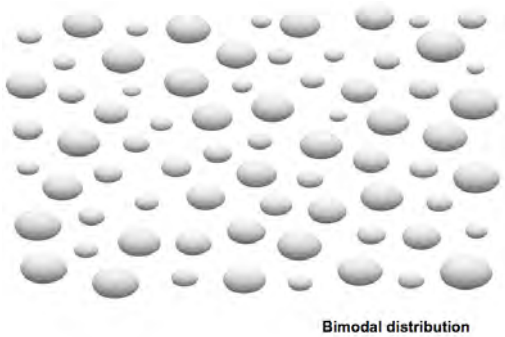


Experimental research on parameter setting





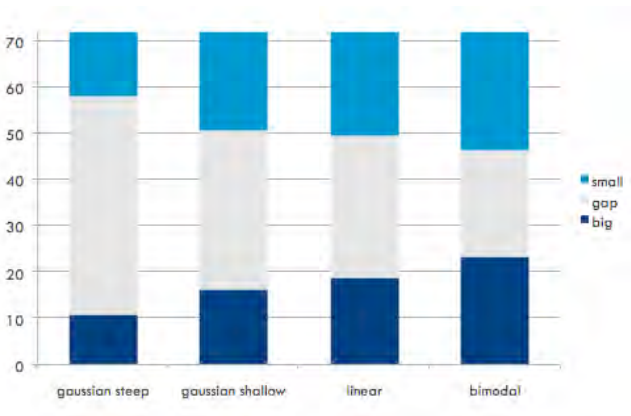


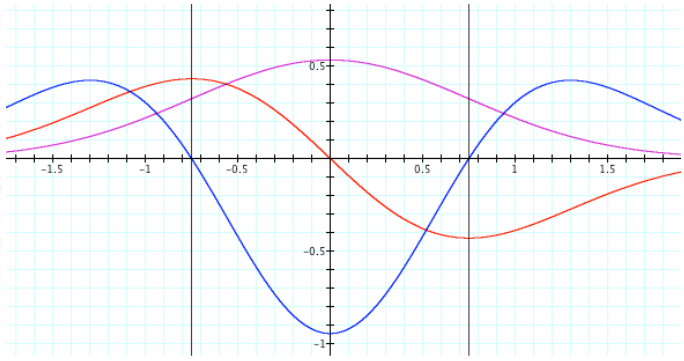


Experimental research on parameter setting

Results:

- The distribution has an influence
- With normal distributions, fewer than 1/3rd of the eggs were assigned to *big* and to *small*, respectively.
- Possible relevance of **standard deviation**:
  the point of inflection of the normal distribution
  where differences between adjacent cohorts are maximized.



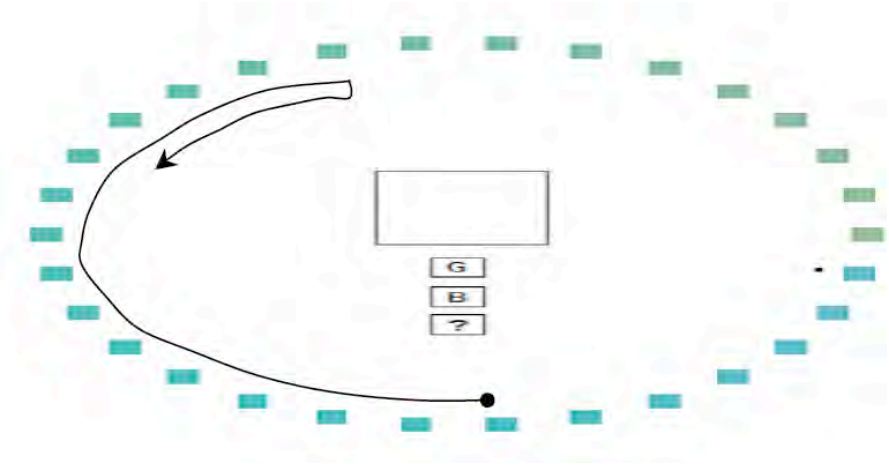


Normal distribution and first and second derivative

Experimental research on parameter setting

Dependence of parameter setting on local comparison class:

- Diana Raffman on hysteresis effects.
- Possible relevance for case-based law:
  which precedence cases are considerd?



Judgement as *blue* or *green* depends on last-seen color spot.

Conclusion

- Apparently precise expressions can be used approximately (number words)
- Apparently vague expressions can be given a precise meaning relative to a parameter that can be set to optimize communication.
- The following assumptions can be entertained:
  - The **literal** meaning of expressions is precise.
  - Vagueness arises as a phenomenon of applying the literal meaning of expressions in communication.
  - Vagueness is a pragmatic phenomenon.