Vagueness & Polysemy 3

Semantic Approaches to Vagueness: Probabilistic Approaches

Peter Sutton
peter.sutton@hhu.de
Heinrich Heine University Düsseldorf

Winter School on Semantics in Barcelona

January, 2018

Main questions for the course

Main Questions

- Why is vagueness a challenge for a formal semantics for natural languages?
  - Day 1: Background
    - Properties of Classical First Order Logic (CFOL) & connection to modern semantic theory
    - Diagnosing Vagueness
- **What are the main reactions to this challenge?**
  - Days 1-3: Responses to vagueness
    - fuzziness (day 1)
    - S'valuationism & polysemy, contextualism, degrees
    - probabilistic approaches
- How does vagueness connect to polysemy?
  - Part of Day 2: Is vagueness a kind of polysemy?
  - Day 4: The common origins of vagueness and polysemy?

Plan for Day 3

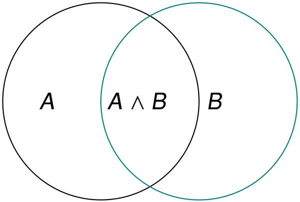
1. Introduction to Probability Theory
   - Axioms and basic theorems
   - Conditional probabilities and Bayesian Updating
2. Verities
3. Probabilistic Linguistic Knowledge
4. Probabilistic Judgements and Correlations

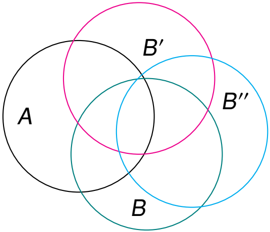
Deductive reasoning vs. Reasoning under uncertainty

Deduction:

May is PM and Brexit will happen ⊨ Brexit will happen ($A \wedge B \models B$)

- Certainty about the premises gives certainty about the conclusion.
- All of the $A \wedge B$ situations are B situations





Reasoning under Uncertainty (Induction):

May is PM $\rightsquigarrow_{defeasible}$ Brexit will happen ($A \rightsquigarrow_{defeasible} B$)

- Certainty about the premises gives *uncertainty* about the conclusion.
- A defeasible inference (based on plausibility, not certainty)
- The number (and plausibility) of B situations that are A situations affects the plausibility of the inference

Axioms of probability

Classical probability theory is a good candidate to model plausibility.
Assume a 'sample space' i.e. a set of all possible outcomes, U (e.g. a set $\{A, B, C....\}$)

- For example, the set of outcomes for a die

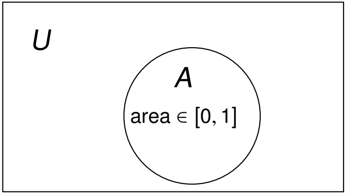
  ⚀ ⚂ ⚅

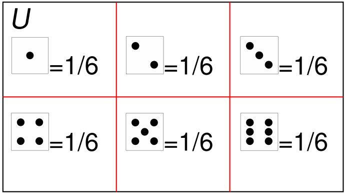
  ⚁ ⚃ ⚄

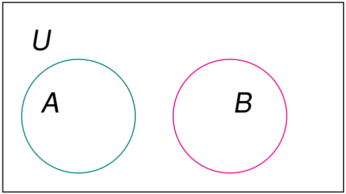
Axioms of Probability

1. A probability measure P such that $P(A) \geq 0$ for any event A
2. $P(U) = 1$
   - $P(\{$⚀, ⚁, ⚂, ⚃, ⚄, ⚅$\}) = 1$
   - There must be an outcome (at least one possibility must be an outcome)
3. If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$

Graphical Representations of Axioms







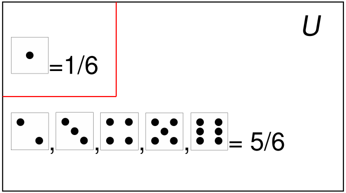
1. A probability measure P such that $P(A) \geq 0$ for any event A
   - Negative areas do not exist
   - $P(A) \in [0, 1]$ for all A

2. $P(U) = 1$
   - The area in $U = 1$

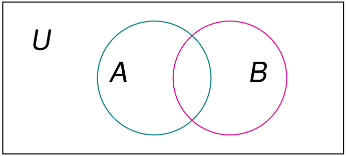
3. If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$
   - Areas are non-overlapping, so just sum the areas

Basic Theorems

1. The empty set has zero probability: $P(\varnothing) = 0$
2. Negation $P(\neg A) = 1 - P(A)$
   in set theoretic notation: $P(\overline{A}) = 1 - P(A)$



3. The addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



When events are not disjoint, minus the overlap to prevent double counting.

(When disjoint $P(A \cap B) = 0$)

Exclusiveness and Independence

Independence: *A* and *B* are independent iff:

$$P(A \cap B) = P(A) \times P(B)$$

Disjoint does *not* mean independent:

Example (for one toss):

Sample space = $\{H, T\}$ (*Heads* or *Tails*)

$P(H) = 0.5$, $P(T) = 0.5$

$P(H \cap T) \neq 0.5 \times 0.5$

Conditional Probability

So far, we have discussed:

- Probabilities of events $P(A)$
- Simple Boolean operations (negation, conjunction, disjunction).
  - Negation/ Compliment: $P(\neg A)/P(\overline{A})$
  - Disjunction/Union: $P(A \vee B)/P(A \cup B)$
  - Conjunction/Intersection: $P(A \wedge B)/P(A \cap B)$

Next up:

- Probability of one event conditional upon another: $P(A|B)$
  - Read this as: *The probability of A, given B*

Bayes' Rule:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Example: Dice roll

$$O = \{\text{⚀}, \text{⚂}, \text{⚄}\}, \qquad M = \{\text{⚁}, \text{⚂}, \text{⚄}\}, \qquad O \cap M = \{\text{⚂}, \text{⚄}\}$$

$$P(O|M) = \frac{P(O \cap M)}{P(M)} = \frac{1/3}{1/2} = 2/3$$

Conditional Probabilities as reduced universes

$$O = \{⚀, ⚂, ⚄\}, \qquad M = \{⚁, ⚂, ⚄\}, \qquad O \cap M = \{⚂, ⚄\}$$

- When we condition on Prime (M), this restricts the universe to only M situations, and *normalizes* the probability space.
  - The new restricted universe: $U_r{=}M$
  - Normalization ensures: $P(U_r{=}M) = 1$

U		
⚀=1/6	⚃=1/6	⚅=1/6
⚁=1/6	⚂=1/6	⚄=1/6

$\mapsto$

$U_r{=}M$		
⚁=1/3	⚂=1/3	⚄=1/3

Then the conditional probabilities are the proportion of the new universe in which $E \cap M$ holds:

$$P(O|M) = \frac{P(O \cap M)}{P(M)} = 2/3$$

$U_r{=}M$	$O \cap M$
	{⚂, ⚄}=2/3

Independence and Conditional Probability

When two events are independent, then conditional probability reduces to unconditional probability:

- $P(A|B) = P(A)$, when A and B are independent

Example: Tossing two coins ($\langle H, T\rangle$ means *coin 1 is heads, coin 2 is tails*)

$$U = \{\langle H, H\rangle, \langle H, T\rangle, \langle T, H\rangle, \langle T, T\rangle\}$$

Suppose:
H1 = Coin 1 is heads
H2 = Coin 2 is heads
The result of each toss is independent, so what is $P(H1|H2)$?

- $P(H1|H2) = \frac{P(H1 \cap H2)}{P(H2)}$

$$\begin{array}{llll} H1 & = \{\langle H, H\rangle, \langle H, T\rangle\} & P(H1) & = 0.5 \\ H2 & = \{\langle H, H\rangle, \langle T, H\rangle\} & P(H2) & = 0.5 \\ H1 \cap H2 & = \{\langle H, H\rangle\} & P(H1 \cap H2) & = 0.25 \end{array}$$

- $P(H1|H2) = \frac{0.25}{0.5} = 0.5 = P(H1)$

Conjunction Rule

We already know how to calculate conjunctions/intersections when A and B are independent:

$$P(A \cap B) = P(B) \times P(A)$$

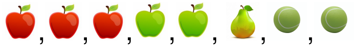
When A and B are not independent, we can just manipulate Bayes' Rule:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

To get the conjunction rule:

$$P(A \cap B) = P(B) \times P(A|B)$$

Example: a bowl contains three red apples, two green apples, one green pear, and two green balls. One is chosen at random.

$U = \{$  $\}$

What is the probability of choosing a green piece of fruit?

$$P(Green \cap Fruit) = P(Green) \times P(Fruit|Green) = 5/8 \times 3/5 = 3/8$$
$$P(Green \cap Fruit) = P(Fruit) \times P(Green|Fruit) = 6/8 \times 1/2 = 3/8$$

Bayes Theorem

We can combine the Conjunction Rule and Bayes' Rule to get Bayes theorem:

$$
\begin{aligned}
P(A|B) &= \frac{P(A \cap B)}{P(B)} \\
P(A \cap B) &= P(A) \times P(B|A) \\
P(A|B) &= \frac{P(A) \times P(B|A)}{P(B)}
\end{aligned}
$$

Some terminology:
$P(A)$, $P(B)$ - The *Priors*
$P(B|A)$ - The *Likelihood*
$P(A|B)$ - The *Posterior*

Applying Bayes' Theorem

In many cases, we do not know, directly, everything to calculate a conditional probability.

Example: Weighing up hypotheses.

Suppose that we have some data, either from a corpus or an experiment. An *n* number of hypotheses about what explains the data. We want to know which hypothesis is most likely.

- The data set D
- The Hypotheses $\{H_1, H_2, ..., H_n\}$

$$P(H_i|D) = \frac{P(H_i) \times P(D|H_i)}{P(D)}$$

- We can estimate $P(H_i)$ either as intuitive plausibility (or as equally distributed).
- We can estimate $P(D|H_i)$ because we know the predictions of our hypotheses.
- BUT we often have no idea how probable the data we have observed is (as opposed to having observed other data).

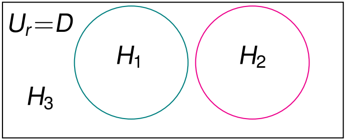
Applying Bayes' Theorem 2

How do we estimate $P(D)$?

$$P(H_i|D) = \frac{P(H_i) \times P(D|H_i)}{P(D)}$$

- Some simple manipulation of Bayes' Theorem can resolve this.
- We can choose hypotheses such that we know that:

$$\sum_{i=1}^{n} P(H_i|D) = 1$$



- E.g. At least one hypothesis must be true. Therefore:

$$1 = \sum_{i=1}^{n} \left(\frac{P(H_i) \times P(D|H_i)}{P(D)} \right) \quad \text{so} \quad P(D) = \sum_{i=1}^{n} P(H_i) \times P(D|H_i)$$

- This gives Bayes' Theorem with $P(D)$ replaced:

$$P(H_i|D) = \frac{P(H_i) \times P(D|H_i)}{\sum_{i=1}^{n} P(H_i) \times P(D|H_i)}$$

Applying Bayes' Theorem: Updating beliefs

Danny lives in a house with a garden and has a cat, Felix. There is a cat-flap for the cat to come and go. Over time, Danny notices that Felix's food consumption is going up significantly, but that Felix is not getting fatter.
$H_i = \{\text{other\_cat\_ate}, \text{felix\_ate}\}$

$E_1 = \text{felix\_not\_fatter}$

$$P(H_i|E_1) = \frac{P(H_i) \times P(E_1|H_i)}{\sum_i P(H_i) \times P(E_1|H_i)}$$

$P(\text{felix\_ate}) = 0.5$ $\quad P(E|\text{felix\_ate}) = 0.25$
$P(\text{other\_cat\_ate}) = 0.5$ $\quad P(E|\text{other\_cat\_ate}) = 0.75$

$$P(\text{felix\_ate}|E_1) = \frac{0.5 \times 0.25}{(0.5 \times 0.25) + (0.5 \times 0.75)} = \frac{0.125}{0.125 + 0.375} = 0.25$$

$$P(\text{other\_cat\_ate}|E_1) = \frac{0.5 \times 0.75}{(0.5 \times 0.25) + (0.5 \times 0.75)} = \frac{0.375}{0.125 + 0.375} = 0.75$$

Updating beliefs cont: Exercise

Felix is a black and white cat. On closer examination, Danny finds some ginger hairs next to the empty cat bowl (looking suspiciously like from the neighbour's cat, Garfield).

$H_i = \{\text{other\_cat\_ate}, \text{felix\_ate}\}$

$E_2 = \text{ginger\_hairs}$

$$P(H_i|E_2) = \frac{P(H_i) \times P(E_2|H_i)}{\sum_i P(H_i) \times P(E_2|H_i)}$$

$P(\text{felix\_ate}) = 0.25$ $\quad$ $P(E_2|\text{felix\_ate}) = 0.25$

$P(\text{other\_cat\_ate}) = 0.75$ $\quad$ $P(E_2|\text{other\_cat\_ate}) = 0.75$

What credences should Danny have now? (easier to do as fractions if you have no calculator)

Updating beliefs cont: Exercise

$H_i = \{\text{other\_cat\_ate}, \text{felix\_ate}\}$

$E_2 = \text{ginger\_hairs}$

$$P(H_i|E_2) = \frac{P(H_i) \times P(E_2|H_i)}{\sum_i P(H_i) \times P(E_2|H_i)}$$

$P(\text{felix\_ate}) = 0.25$	$P(E_2|\text{felix\_ate}) = 0.25$
$P(\text{other\_cat\_ate}) = 0.75$	$P(E_2|\text{other\_cat\_ate}) = 0.75$

What credences should Danny have now? (easier to do as fractions if you have no calculator)

	$P(\text{felix\_ate}|E_2)$	$P(\text{other\_cat\_ate}|E_2)$
$P(H_i) \times P(E_2|H_i$	1/4 x 1/4 = 1/16	3/4 x 3/4 = 9/16
$\sum_i P(H_i) \times P(E_2|H_i$	1/16 + 9/16 = 10/16	1/16 + 9/16 = 10/16
$P(H_i|E_2)$	1/16 x 16/10 = 1/10	9/16 x 16/10 = 9/10

Verities: Degrees of closeness to clear truth

- Edgington (1992, 1997) argues that a logic for vagueness shares a structural similarity with classical Bayesian probability calculus.
  - Main goal: a logic that is better than a fuzzy logic
- Verities: degrees of closeness to clear cases of truth
  - In range $[0, 1]$
  - Not degrees of (un)certainty

Interpretation of Propositions as Verities

1. $\mathcal{I}_v(\phi) \in [0, 1]$
2. $\mathcal{I}_v(\neg\phi) = 1 - \mathcal{I}_v(\phi)$
3. $\mathcal{I}_v(\phi \wedge \psi) = \mathcal{I}_v(\phi) \times \mathcal{I}_v(\psi|\phi)$
4. $\mathcal{I}_v(\phi \vee \psi) = \mathcal{I}_v(\phi) + \mathcal{I}_v(\psi) - \mathcal{I}_v(\phi \wedge \psi)$

Verities: Degrees, but not degree functionality

Conditional verities i.e., $\mathcal{I}_v(\phi|\psi)$ mean no degree functionality

- Fuzzy logic: degree of truth for $\phi X \psi$, for any connective X is a function of the degrees of truth of ϕ and ψ
- Verities: No degree functionality

Example:

- Alex is 190cm in height
  - Suppose: $\mathcal{I}_v(\mathit{Tall}(a)) = 0.9$
- Billie is 185cm in height
  - Suppose: $\mathcal{I}_v(\mathit{Tall}(b)) = 0.7$
- What is the verity for $\mathit{Tall}(a) \wedge \mathit{Tall}(b)$?
  - b is shorter that a, so $\mathcal{I}_v(\mathit{Tall}(a)|\mathit{Tall}(b)) = 1$
  - $\mathcal{I}_v(\mathit{Tall}(a) \wedge \mathit{Tall}(b)) = \mathcal{I}_v(\mathit{Tall}(b)) \times \mathcal{I}_v(\mathit{Tall}(a)|\mathit{Tall}(b)) = 0.5 \times 1 = 0.7$

Edgington's argument against Fuzzy views (from Day 1)

Suppose that Alex is borderline tall

- $\mathcal{I}_f(\mathit{Tall}(a)) = 0.5$

Also suppose that Billie is marginally less tall than Alex, e.g.

- $\mathcal{I}_f(\mathit{Tall}(b)) = 0.4$

Shouldn't it be perfectly false to say that Billie is tall, but Alex is not?

Not with fuzzy connectives

- $\mathcal{I}_f(\mathit{Tall}(b) \wedge \neg \mathit{Tall}(a)) = \mathit{min}(\{0.4, 1 - 0.5\} = \mathit{min}\{0.4, 0.5\} = 0.4$

Verities can handle this case

Suppose that Alex is borderline tall

- $\mathcal{I}_v(\mathit{Tall}(a)) = 0.5$

Also suppose that Billie is marginally less tall than Alex, e.g.

- $\mathcal{I}_v(\mathit{Tall}(b)) = 0.4$

Shouldn't it be perfectly false to say that Billie is tall, but Alex is not?

Yes!
$\mathcal{I}_v(\neg \mathit{Tall}(a)) = 1 - \mathcal{I}_v(\mathit{Tall}(a)) = 0.5$
$\mathcal{I}_v(\mathit{Tall}(b) | \neg \mathit{Tall}(a)) = 0$

$$\begin{aligned} \mathcal{I}_v(\mathit{Tall}(b) \wedge \neg \mathit{Tall}(a)) &= \mathcal{I}_v(\neg \mathit{Tall}(a)) \times \mathcal{I}_v(\mathit{Tall}(b) | \neg \mathit{Tall}(a)) \\ &= 0.5 \times 0 \\ &= 0 \end{aligned}$$

Verities: Logical Consequence I

- Like supervaluationism, a verity-based account also results in a divergence from classical consequence for multi-premise conclusions.
- Recall the disjunction rule:
  - $\mathcal{I}_v(\phi \vee \psi) = \mathcal{I}_v(\phi) + \mathcal{I}_v(\psi) - \mathcal{I}_v(\phi \wedge \psi)$
- So it is possible that
  - $\mathcal{I}_v(\phi \vee \psi) = 1$ when $\mathcal{I}_v(\phi) < 1$ and $\mathcal{I}_v(\psi) < 1$

Verities: Logical Consequence II

Edgington's book example:

"A library book can be such that it is not clear whether it should be classified as Philosophy of Language or Philosophy of Logic; but if we have a joint category for books of either kind, it clearly belongs there."

Suppose, for book b:

$$\begin{array}{lcl} \mathcal{I}_v(Lang(b)) & = & 0.6 \\ \mathcal{I}_v(Logic(b)) & = & 0.6 \\ \mathcal{I}_v(Logic(b)|Lang(b)) & = & \mathcal{I}_v(Lang(b)|Logic(b)) = 0.\overline{3} \end{array}$$

- So Phil. of Lang and Logic books overlap in some way.

$$\begin{array}{lcl} \mathcal{I}_v(Logic(b) \vee Lang(b)) & = & \mathcal{I}_v(Lang(b)) + \mathcal{I}_v(Logic(b)) \\ & & -\mathcal{I}_v(Logic(b) \wedge Lang(b)) \\ & = & (0.6 + 0.6) - (0.6 \times 0.\overline{3}) \\ & = & 1 \end{array}$$

Verities: Logical Consequence III

Furthermore, verities preserve LEM and non-contradiction, but do not support multi-conclusions for arbitrary propositions and their negations

- In other words, like supervaluationism, we have weak paraconsistency:

$$\psi \vDash_{\mathrm{v}} \phi \vee \neg\phi$$
$$\psi \nvDash_{\mathrm{v}} \phi, \neg\phi$$

However, for the multi-premise, single conclusion case:

- For $P1, ..., P_n \nvDash_{\mathrm{v}} C$
  - $1 - (\sum_i \mathcal{I}_v(P_i)) \leq 1 - \mathcal{I}_v(C)$
  - The sum of the unverity of the premises is equal to the unverity of the conclusion
  - See Edgington (1997, pp. 307-8) for the proof

Verities and the Sorites I

Assume the following verities

$$\mathcal{I}_v(Tall(a_1)) = 1 \quad \mathcal{I}_v(Tall(a_2)) = 0.8 \quad \mathcal{I}_v(Tall(a_3)) = 0.6$$
$$\mathcal{I}_v(Tall(a_4)) = 0.4 \quad \mathcal{I}_v(Tall(a_5)) = 0.2 \quad \mathcal{I}_v(Tall(a_6)) = 0$$

A long sorites argument:

(P1) $Tall(a_1)$

(P2) $Tall(a_1) \rightarrow Tall(a_2)$

(P3) $Tall(a_2) \rightarrow Tall(a_3)$

(P4) $Tall(a_3) \rightarrow Tall(a_4)$

(P5) $Tall(a_4) \rightarrow Tall(a_5)$

(P6) $Tall(a_5) \rightarrow Tall(a_6)$

(C) $Tall(a_6)$

Provably:

- The unverity for each material conditional is 0.2
- So, the unverity of the conclusion must be ≥ 1
- So, if valid, the conclusion must have a verity of 0

The argument is valid, but unsound, since the sum of the unverities of the premises is 1 (the premises taken together are clearly false)

Verities and the Sorites II

However, like with superavaluationism, we get bad results for the short sorites:

A short sorites argument:

(P1) $Tall(a_1)$

(P2) $\forall x_\forall x_{i+1}[Tall(x_i) \rightarrow Tall(x_{i+1})]$

(C) $Tall(a_6)$

- (P1) has verity 1
- (P2) and (C) have verity 0

The argument is still valid, but unsound, since P2 is clearly false.

However:

- If $\forall x_\forall x_{i+1}[Tall(x_i) \rightarrow Tall(x_{i+1})]$ has verity 0, then $\exists x \exists x_{i+1}[Tall(x_i) \wedge \neg Tall(x_{i+1})]$ has verity 1

But that means that *Tall* has sharp boundaries. Is it still vague?

- However, perhaps not as bad a supervaluationism
  - All this reflects is that not such conjunct is clearly true
  - Forced to think of $\exists x\phi$ as a long disjunct
  - Every instance ϕ_i is almost clearly true

Objections against verities

- Like with supervaluationism, we can object to the affect on disjunction

$$\psi \vDash_{\mathsf{V}} \phi \vee \neg\phi$$
$$\psi \nvDash_{\mathsf{V}} \phi, \neg\phi$$

- Likewise to the falsity of tolerance statements
- Strings of conjunctions
  - An advantage of fuzzy systems: Suppose the following are all borderline true: Alex is tall, Billie is bald, Charlie is slim
  - Their conjunction is also borderline true.
  - Not with verities: $\mathcal{I}_v$(Alex is tall $\wedge$ Billie is bald $\wedge$ Charlie is slim) = 0.125
  - Is this a problem though?
    - With every new conjunct, there is a greater risk of going wrong
  - A sweeter pill to swallow than the fuzzy results for dependent sentences e.g., *Billie is tall and Alex is not tall* (when Alex is taller than Billie)

Are verities just probabilities?

Edgington thinks not:

Case 1: Probability.

- Suppose I want coffee. Tea would do, but is less preferable. I can go to ask three colleagues, P1, P2, and P3 for a cup:
  - (P1) always has coffee
  - (P2) has coffee 50% of the time and tea 50% of the time
  - (P3) always has tea
- My clear order of preference is P1, P2, P3

Case 2: Verities.

- Suppose I want coffee. Tea would do, but is less preferable. I can go to ask three colleagues, V1, V2, and V3 for a cup:
  - (V1) always has coffee
  - (V2) is a bit of a dreamer, and usually ends up making coffee in a pot still half filled with tea leaving some kind of tea/coffee borderline case
  - (V3) always has tea
- Now my clear order of preference starts with P1
- However, my preference for coffee over tea does not imply a preference for borderline coffee/borderline tea over tea

Why verities might be metalinguistic probabilities

Edgington's argument works by comparing verities with uncertainty about the world

- A more apt comparison is uncertainty about what the expressions mean (Sutton 2013)

Case 3: Metalinguistic Probability.

- Suppose I want coffee. Tea would do, but is less preferable. I can go to ask three colleagues, M1, M2, and M3 for a cup:

  (M1) always makes something that I am certain *coffee* describes
  (M2) is a bit of a dreamer, and usually ends up making coffee in a pot still half filled with tea leaving something such that I am uncertain whether *coffee* decribes or *tea* describes
  (M3) always makes something that I am certain *tea* describes

- Now my clear order of preference starts with P1
- However, my preference for coffee over tea does not imply a preference for borderline coffee/borderline tea over tea

Metalinguistic uncertainty aligns with verities.

More worldly uncertainty and metalinguistic uncertainty

Barker (2002) raised this distinction (I think first in the vagueness literature)

- Suppose it is uncertain what the standards for *tall* are in the context.
- You are told 'Alex is tall'. This allows you to
  (i) Update the common ground with $Tall(a)$
  (ii) Update the standard of tallness in the context (If you know roughly Alex's height)
    - ★ If Alex's hight is h, you learn that (in the context) $\forall x[Height(x) \geq h \rightarrow Tall(x)]$

Lassiter (2011) suggests a dynamic Bayesian model for both kinds of update.

Vagueness as probabilistic linguistic knowledge

- At core epistemicist but reduces the implausibility burden
  - Uncertainty about exact standards in the context, not about the meaning of vague expressions per se
- As such, a communication centric position
  - Agents communicate and narrow down meanings of vague expressions sufficiently for the purposes at hand
- Leads to a kind of probabilistic supervaluationism with an epistemicist flavour

Probabilistic Linguistic Knowledge

Uncertainty about:

- Which possible world is the actual world
- Which possible (precise) language is being used in the context

Agents reason about boundaries for vague expressions given uncertainty about how the world is

- Vagueness is a side effect of rarely (if ever) completely removing all this uncertainty

Example

Suppose we do not know Alex's height

- Possible options as possible worlds, $W = \{w_1, ..., w_n\}$

$$w_1 \supseteq \{Height(a) = 158cm\}$$
$$w_2 \supseteq \{Height(a) = 188cm\}$$

We also do not know the standard for height in the context:

- Possible options modelled as possible languages/precisifications, $L = \{l_1, ..., l_n\}$

$$l_1 = \lambda x.height(x) \geq 150cm; \quad l_2 = \lambda x.height(x) \geq 160cm;$$
$$l_3 = \lambda x.height(x) \geq 170cm; \quad l_4 = \lambda x.height(x) \geq 180cm;$$
$$l_5 = \lambda x.height(x) \geq 190cm$$

Uncertainty modelled as a function $\mu : \langle W \times L \rangle \rightarrow [0, 1]$

$$\mu = \{$$
$$\langle\langle w_1, l_1\rangle, 0.02\rangle, \langle\langle w_1, l_2\rangle, 0.08\rangle, \langle\langle w_1, l_3\rangle, 0.15\rangle, \langle\langle w_1, l_4\rangle, 0.2\rangle, \langle\langle w_1, l_5\rangle, 0.05\rangle,$$
$$\langle\langle w_2, l_1\rangle, 0.02\rangle, \langle\langle w_2, l_2\rangle, 0.08\rangle, \langle\langle w_2, l_3\rangle, 0.15\rangle, \langle\langle w_2, l_4\rangle, 0.2\rangle, \langle\langle w_2, l_5\rangle, 0.05\rangle$$
$$\}$$

Example cont.

$$\mu = \{ \langle\langle w_1, l_1\rangle, 0.02\rangle, \langle\langle w_1, l_2\rangle, 0.08\rangle, \langle\langle w_1, l_3\rangle, 0.15\rangle, \langle\langle w_1, l_4\rangle, 0.2\rangle, \langle\langle w_1, l_5\rangle, 0.05\rangle, \langle\langle w_2, l_1\rangle, 0.02\rangle, \langle\langle w_2, l_2\rangle, 0.08\rangle, \langle\langle w_2, l_3\rangle, 0.15\rangle, \langle\langle w_2, l_4\rangle, 0.2\rangle, \langle\langle w_2, l_5\rangle, 0.05\rangle \}$$

We can calculate probabilities of possible worlds:

$$\mu_W(w') = \sum_{l \in L} \mu(w', l)$$

$$\mu_W(w_1) = \mu_W(w_2) = 0.02 + 0.08 + 0.15 + 0.2 + 0.05 = 0.5$$

We can also calculate probabilities of possible languages:

$$\mu_L(l') = \sum_{w \in W} \mu(w, l')$$

$$\mu_L(l_1) = 0.02 + 0.02 = 0.04 \quad \mu_L(l_2) = 0.08 + 0.08 = 0.16$$
$$\mu_L(l_3) = 0.15 + 0.15 = 0.3 \quad \mu_L(l_4) = 0.2 + 0.2 = 0.4$$
$$\mu_L(l_5) = 0.05 + 0.05 = 0.1$$

Example: probabilities of propositions

$$w_1 \supseteq \{Height(a) = 158cm\}$$
$$w_2 \supseteq \{Height(a) = 188cm\}$$

$$l_1 = \lambda x.height(x) \geq 150cm; \quad l_2 = \lambda x.height(x) \geq 160cm;$$
$$l_3 = \lambda x.height(x) \geq 170cm; \quad l_4 = \lambda x.height(x) \geq 180cm;$$
$$l_5 = \lambda x.height(x) \geq 190cm$$

$$\begin{array}{l|ll} \mu_W(w_1) = 0.5 & \mu_L(l_1) = 0.04 & \mu_L(l_2) = 0.16 \\ \mu_W(w_2) = 0.5 & \mu_L(l_3) = 0.3 & \mu_L(l_4) = 0.4 \\ & \mu_L(l_5) = 0.1 & \end{array}$$

Now we can calculate probabilities of propositions:

$$P(\phi) = \sum_{w_i} \sum_{l_i} \mu_W(w_i) \times \mu_L(l_i : True(\phi(w_i)))$$

The probability of the proposition *Tall*(a):

$$\begin{aligned} P(Tall(a)) &= \quad (0.5 \times 0.04) + \\ &\quad\;\; (0.5 \times 0.04) + (0.5 \times 0.16) + (0.5 \times 0.3) + (0.5 \times 0.4) \\ &= 0.02 + 0.45 \\ &= 0.47 \end{aligned}$$

Example: updating probabilities

$$w_1 \supseteq \{Height(a) = 158cm\}$$
$$w_2 \supseteq \{Height(a) = 188cm\}$$

$$l_1 = \lambda x.height(x) \geq 150cm; \quad l_2 = \lambda x.height(x) \geq 160cm;$$
$$l_3 = \lambda x.height(x) \geq 170cm; \quad l_4 = \lambda x.height(x) \geq 180cm;$$
$$l_5 = \lambda x.height(x) \geq 190cm$$

$$\begin{array}{l|ll} \mu_W(w_1) = 0.5 & \mu_L(l_1) = 0.04 & \mu_L(l_2) = 0.16 \\ \mu_W(w_2) = 0.5 & \mu_L(l_3) = 0.3 & \mu_L(l_4) = 0.4 \\ & \mu_L(l_5) = 0.1 & \end{array}$$

But this can be updated.

- We learn that Alex is 158cm (now only w_1 is possible)

$$\begin{aligned} P(Tall(a)|w_1) &= (1 \times 0.04) \\ &= 0.04 \end{aligned}$$

- We learn that Alex is 188cm (now only w_2 is possible)

$$\begin{aligned} P(Tall(a)|w_2) &= (1 \times 0.04) + (1 \times 0.16) + (1 \times 0.3) + (1 \times 0.4) \\ &= 0.9 \end{aligned}$$

PLK and the Sorites

Tolerance principles need to be clarified

$$\text{For all } a, b, \text{if } P(a) \text{ and } a \sim_P b, \text{ then } P(b)$$

Does not take into account that propositions are probabilistic. The following is unproblematically false:

$$\text{For all } a, b, \text{if } p(P(a)) = 1 \text{ and } a \sim_P b, \text{ then } p(P(b)) = 0$$

We can accept the following tolerance principle

$$\text{For all } a, b, \text{if } p(P(a)) = n \text{ and } a \sim_P b, \text{ then } p(P(b)) = n \pm \epsilon$$

Summary and Criticisms of PLK

What is PLK?

- A probabilistic epistemicism/supervaluationism hybrid

Some progress:

- Unlike supervaluationism: No sharp jumps between clearly ϕ and not clearly ϕ
- Unlike epistemicism: No mysterious ignorance about meanings
  - Instead: reasoning about speaker meaning, given uncertainty about contextual standards and what the world is like

Some worries:

- Like supervaluationism: unclear what the cognitive status of precisifications are
- Like epistemicism: There is a *de facto* sharp boundary in every context

PLK and epistemicism

- Probabilities of sentences are probabilities sentences being true
  - For any context, there is a fact of the matter whether e.g. Alex is tall or not
- Even if these probabilities are subjective, the agent is assuming that there is *a* standard for e.g., *tall* in every context

An alternative?

- We do not reason about what exact truth-conditions are
- Instead: meanings are inherently graded
- The cost: meanings do not determine truth-conditions

Getting to the alternative I

A version of Wright's puzzle about tolerance (in terms of competancy):

(P1) *There are fully competent speakers of languages.*
(P2) *If a speaker, S, is fully competent with a predicate F, then S knows under what conditions F would be truly applied.*
(P3) *For vague predicates, no speakers know (entirely/exactly) under what conditions F would be truly applied.*

(C1) *No speakers are fully competent.*

Getting to the alternative II

Epistemically oriented approaches adjust (P2) and (P3)

(P1) *There are fully competent speakers of languages.*

(P2*) *If a speaker, S, is fully competent with a predicate F, then S knows under what conditions F would be truly applied, but only within a margin for error/only within certain accuracy of the contextual standards.*

(P3*) *For vague predicates, no speakers know under what conditions F would be truly applied without any margin for error/without any uncertainty about the contextual standards.*

Getting to the alternative II

There is another option. Take the absurdity of (C) to indicate the falsity of (P2)

(P1) *There are fully competent speakers of languages.*
(P2) *If a speaker, S, is fully competent with a predicate F, then S knows under what conditions F would be truly applied.*
(P3) *For vague predicates, no speakers know (entirely/exactly) under what conditions F would be truly applied.*

(C1) *No speakers are fully competent.*
(C2) *(P2) is false*
(C3) *Speakers can be fully competent without knowing the truth conditions for the predicates they are competent with.*

The difference with an epistemicist account:

- Imperfectly knowing something precise
- Perfectly knowing something imprecise

Probabilistic Judgements and Correlations models

Top-down vs. Bottom-up probabilistic models (Terms coined by van Eijck and Lappin 2012)

- Top-down: Probability distribution over e.g. worlds, sharp languages
  - updated given new information
  - PLK
- Bottom-up: Probability distributions calculated from e.g. situations
  - Learning correlations between types
  - Forming probabilistic type judgements
  - Cooper et al. (2015); Sutton (2015)

Probabilistic Judgements Models: Meaning represented as probability distributions

Probability distributions are learnt and dynamically updated from situations of use

- E.g., $P(\mathit{Tall}(a)|\mathit{Height}(a) = h)$ is calculated from correlations between *Tall* judgements and witnessed heights.
- Need not be truth functional
  - Agent can be highly certain that $\mathit{Tall}(a)$, and still be wrong.

Other PJC accounts

- Some very early probabilistic approaches: Borel (1907/2014), Black (1937)
- Hampton (2007): The treatment of vague concepts via prototypes (represented, more or less, in probabilistic terms)
- Decock and Douven (2014): A conceptual spaces rendering of Hampton's ideas
- Some of my own work: Sutton (2013, 2015, 2017)

Sutton's (2015) version of PJC adopts the distinction made in Situation Semantics between a *Discourse Situation Type* and a *Described Situation Type*.

$$\lambda s(s \vDash \langle\langle \mathtt{height}{=}180\mathtt{cm}, \mathrm{a}, \mathit{yes}\rangle\rangle)$$

- The type of situation in which Alex has a height of 180cm

$$\lambda x \lambda s(s \vDash \langle\langle \mathtt{asserts}, x, \mathrm{Tall}, \mathrm{a}, \mathit{yes}\rangle\rangle)$$

- The type of situation in which some agent calls Alex 'tall'.

Mastering 'tall' amounts to tracking correlations such as:

$$P(\lambda x(\lambda h \lambda s(s \vDash \langle\langle \mathtt{height}{=}\mathtt{hcm}, x, \mathit{yes}\rangle\rangle) \mid \lambda y \lambda s(s \vDash \langle\langle \mathtt{asserts}, y, \mathrm{Tall}, x, \mathit{yes}\rangle\rangle)))$$

- I.e. How do assertions of individuals being Tall correlate with their heights

We'll use a simpler notation:

- $P(s : \mathit{Tall}(a) | s : \mathit{Height}(a) = h)$
  - Important to remember that this does not mean the probability that *a* is tall, given her height
  - It means the probability of an agent judging *a* to be tall, given her height

Example: Judgement sets, calculating priors & liklihoods

Learners record judgements witnessed (can contain noisy or inaccurate data)

Situation	Height (cm)	Judged Tall
s_1	170	*yes*
s_2	182	*yes*
s_3	180	*yes*
s_4	184	*yes*
s_5	176	*yes*
s_6	188	*yes*
s_7	190	*yes*
s_8	189	*yes*
s_9	195	*yes*
s_{10}	191	*yes*

Plus 10 situations with negative Tall judgements

These are used to calculate posteriors and priors, e.g.

height in cm (h)	$h < 175cm$	$175 \leq h < 185$	$h \geq 185$
$P(s : Height(x) = h \mid s : Tall(x))$	0.1	0.4	0.5

height in cm (h)	$h < 175cm$	$175 \leq h < 185$	$h \geq 185$
$P(s : Height(x) = h)$	0.25	0.5	0.25

	Yes	No
$P(s : Tall(x))$	0.5	0.5

Example: Calculating Posteriors

height in cm (h)	$h < 175cm$	$175 \leq h < 185$	$h \geq 185$
$P(s : Height(x) = h \mid s : Tall(x))$	0.1	0.4	0.5

height in cm (h)	$h < 175cm$	$175 \leq h < 185$	$h \geq 185$
$P(s : Height(x) = h)$	0.25	0.5	0.25

	Yes	No
$P(s : Tall(x))$	0.5	0.5

- Application of Bayes' rule

height in cm (h)	$h < 175cm$	$175 \leq h < 185$	$h \geq 185$
$P(s : Tall(x) \mid s : Height(x) = h)$	0.2	0.4	1

- Such distributions *are* the meaning representations
  - No commitment to truth conditions

Bottom-up approaches and the sorites

Do bottom-up probabilistic approaches get us any further?

$\mathtt{h}$	160	170	180	190
$p(S \models Tall(x) \mid S \models x \approx \mathtt{h}cm)$	0.1	0.2	0.5	0.9

These probability values would indicate the probability of a competent speaker using *tall* versus *not tall*.

- No sharp cut-off point
- Truth-conditions are *not* determined by meaning.
- Small differences in height lead to small differences in probability values

BUT: What about correctness conditions?

- When would it be (overall) correct to use *tall* rather than *not tall*?
- $>$ 0.5, $>$ 0.6, $>$ 0.7...?

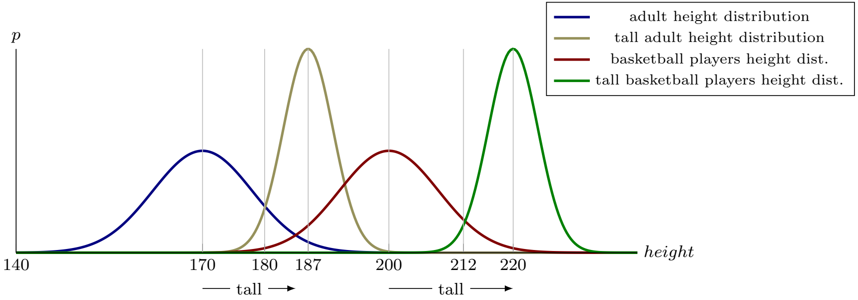
The *deep problem* of vagueness (See Wright 1975; Sutton 2017):

- Removing sharp boundaries at one level makes them re-emerge at another level.
- Proposed ‘solution’ in Sutton (2017)

Bottom-up probabilistic semantics and context and comparison classes

So far, we assumed that adjectives denote predicates.

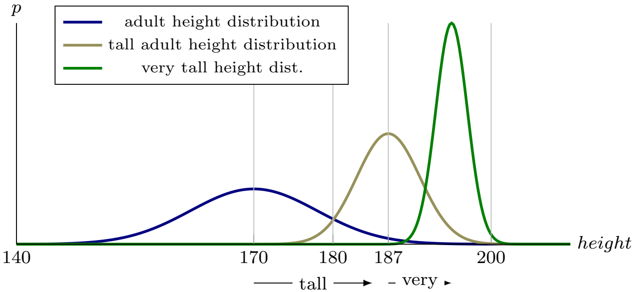
- Sutton (2013) argues that gradable adjectives (e.g. *tall*) can be modelled as functions on probability distributions.
- Gloss: *tall* modifies expectations of height upwards relative to the comparison class



- Borderline case for e.g *tall*: crossing point between *tall* and *not tall*

A natural extension to modifiers?

If gradable denote functions on distributions, it is natural to think of modifiers like *very* doing the same thing:

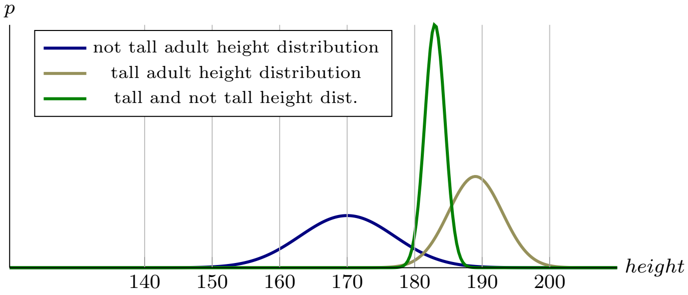


Some criticisms of probabilistic judgements approaches

- Not exactly semantics
  - Lewis (1970): semantics without truth-conditions is not semantics
  - However, meaning before truth-conditions approaches have a long history (Grice, Austin)
  - Nonetheless, a complete collapse of the traditional pragmatics/semantics divide
- Holism
  - If meanings are grounded in correlations, isn't every meaning representation connected to every other, no matter how tangentially?
  - More of a worry, but some restrictions could be implemented
  - Anyway a worry for pragmatic accounts that rely on background knowledge/beliefs
- Truth question left unanswered?
  - Is there a final point at which it is true to say $P(a_i)$ in a sorites series?

Some progress nonetheless?

- The tolerance conditional gets a low/zero probability
  - Harmless: Just means that there is a point at which a competent agent will switch their judgement
- Some possibility to capture assertions of contradictions?
  - Combining probability distributions



Higher Order Vagueness

Maybe a better solution here...

- The challenge
  - At what level of certainty is it correct/justified to assert that ϕ?
- One possible answer (Sutton 2017)
  - The fact that we have dropped meaning in terms of truth conditions means that we can be fully competent speakers, know the meanings of expressions, but still not know where in the series to stop.
  - Relocate the sharp boundary:
    - ★ There might be a correct place to stop asserting ϕ, but if it is in a place we should not be expected to know, then all is well
    - ★ Proposal: Externalise correctness to communicative burden–have we done enough to communicate the intended message to our interlocutor in the situation?
    - ★ This standard of correctness is not something we can be expected to know
    - ★ Sharp boundaries relocated to an innocuous position
- This is subtly different from epistemicist approaches where there is an unknown boundary determined by the meaning of an expression (in context)

Summary

	Fuzzy Logic	Superval	Subval	Kamp's Contextualism	Epistemicism/ Degrees/ Verities/PLK
No gaps	No	No	Yes	Yes	Yes
No gluts	Yes	Yes	No	Yes	Yes
Excluded Middle	No	Yes	Yes	Yes	Yes
Non-Contradiction	No	Yes	Yes	Yes	Yes
Modus Ponens	Yes	Yes	No	Yes	Yes
Quantifier Duality	Yes	Yes	Yes	No	Yes

PJC sits a bit outside of this classification, since it does not commit to truth functionality

Take-home message for day 3

- Verities capture dependencies between propositions (like supervaluationism)
  - Seems to end up being paracomplete (like supervaluationism)
  - Has short sorites problem (like supervaluationism)
  - But a more plausible solution: No big jumps between antecedents and consequents
- PLK gives us a model for reasoning about the standards for terms, given contextual worldly information and metasemantic information
  - A sort of epistemicism-supervaluationism hybrid
  - Makes the epistemicist conclusions more palatable?
  - Replaces tolerance with a clarification:
    For all $a, b,$ if $p(P(a)) = n$ and $a \sim_P b,$ then $p(P(b)) = n \pm \epsilon$
  - Still leads to HOV problems

Take-home message for day 3

- PJC also gives us a model for reasoning about the standards for terms, given contextual worldly information and metasemantic information
  - Also assumes no top-down probability distribution over worlds/languages
  - Rather radical: does not adopt a truth-conditional semantics
  - Less of a short sorites problem
  - But a more plausible solution: No big jumps between antecedents and consequents
  - Arguably has a more plausible treatment of

Thanks

Especial thanks to Robin Cooper, Paul Egré, and Shalom Lappin for many helpful comments and conversations.

Selected References I

Chris Barker. The dynamics of vagueness. *Linguistics and Philosophy*, 25(1):1–36, 2002.

Max Black. Vagueness. an exercise in logical analysis. *Philosophy of Science*, 4(4):427–455, 1937.

Émile Borel. An economic paradox: The sophism of the heap of wheat and statistical truths. *Erkenntnis*, 79(Suppl. 5):1081–1088, 1907/2014. ISSN 1572-8420. doi: 10.1007/s10670-014-9615-z. URL `http://dx.doi.org/10.1007/s10670-014-9615-z`.

Robin Cooper, Simon Dobnik, Shalom Lappin, and Staffan Larsson. Probabilistic type theory and natural language semantics. *Linguistic issues in Language Technology*, 10, 2015.

Lieven Decock and Igor Douven. What is graded membership? *Noûs*, 48 (4):653–682, 2014.

Dorothy Edgington. Validity, uncertainty and vagueness. *Analysis*, 52(4): 193–204, 1992.

Selected References II

Dorothy Edgington. Vagueness by degrees. In R. Keefe and P. Smith, editors, *Vagueness: A Reader*, pages 294–316. MIT Press, Cambridge, MA, 1997.

James A Hampton. Typicality, graded membership, and vagueness. *Cognitive Science*, 31(3):355–384, 2007.

Daniel Lassiter. Vagueness as probabilistic linguistic knowledge. In R. Nouwen, U. Sauerland, H.C. Schmitz, and R. van Rooij, editors, *Vagueness in Communication*, pages 127–150. Springer, 2011.

David Lewis. General semantics. *Synthese*, 22:18–67, 1970.

Peter Sutton. *Vagueness, Communication, and Semantic Information*. PhD thesis, King's College London, 2013.

Peter R. Sutton. Towards a probabilistic semantics for vague adjectives. In Hans-Christian Schmitz and Henk Zeevat, editors, *Language, Cognition, and Mind*. Springer, 2015.

Selected References III

Peter R. Sutton. Probabilistic approaches to vagueness and semantic competency. Erkenntnis: published as an online first article, 2017.

J. van Eijck and S. Lappin. Probabilistic semantics for natural language. In Z. Christoff, P. Galeazzi, N. Gierasimszuk, A. Marcoci, and S. Smets, editors, *Logic and Interactive Rationality (LIRA) 2012, Volume 2*, pages 17–35. ILLC, University of Amsterdams, 2012.

Crispin Wright. On the coherence of vague predicates. *Synthese*, 30 (3-4):325–365, 1975.