Probabilistic Approaches to Semantics

Ariel Cohen

Department of Foreign Literatures and Linguistics
Ben-Gurion University of the Negev
arikc@bgumail.bgu.ac.il

The goals of this course

- Probability is a tool, just like logic or algebra.
- Recently, there has been considerable interest in the applicability of probability to linguistics.
- Our goal: to discuss the applicability of this tool to semantics, its strengths and weaknesses.
- We will consider a number of test cases where probability is applied to a semantic problem.
- For each case, we will ask ourselves:
  - Is a probabilistic treatment appropriate?
  - If so, what type of probabilistic account should we use?
  - What are its advantages and disadvantages?
- Hopefully: we will end up with another weapon in our arsenal as we attack semantic problems.

The Mathematics of Probability

- The axioms of probability
- Conditional probability
- Bayes's theorem
- Random variables
- Expectation

The Sample Space

- Probability only makes sense in the context of a particular set of possible outcomes.
- The sample space Ω.
- Cast a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$
- Any proposition is a subset of Ω
- "The die came up 3" is $\{3\}$
- "The die came up an even number" is $\{2, 4, 6\}$.
- $\ldots$

The axioms

Any function that satisfies the following (slightly simplified) axioms is a probability function:

- For all $A \subseteq \Omega$, $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- Countable additivity: for any disjoint sets $A_1, A_2. \ldots \subseteq \Omega$,

  (1) $P(\cup A_i) = \sum_i P(A_i)$

Some consequences

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(\emptyset) = 0$
- $P(\overline{A}) = 1 - P(A)$
- If $A \subseteq B$ then $P(A) \leq P(B)$
- $P(B - A) = P(B) - P(A \cap B)$

Conditional probability

- Usually we want not the probability that A happens, but the probability that A happens given that B happens.
- For example: we want the probability that if something is a bird, then it flies, not the probability that something flies.
- Conditional probability:

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

- Two events are independent if

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

- Equivalently:

(4) $P(A|B) = P(A)$

Bayes's theorem

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

Random variables

- A random variable is a function (not a variable!) from the sample sapce Ω to real numbers.
- For any value x of the random variable X, we can talk about the set of events that cause this value:

  (6) $A_x = \{\omega \in \Omega : X(\omega) = x\}$

- The probability mass function:

  (7) $p(x) = p(X = x) = P(A_x)$

Expectation

The expectation is the average of a random variable:

(8) $E(X) = \sum_x (x \times p(x))$

Probability and semantics

- Mathematical probability theory tells us how to calculate probabilities, but not what they mean.
- What are the truth conditions of $P(A|B) = 0.8$?
- Thus, probability judgments themselves need a semantics.

Generics and frequency adverbs

- The phenomenon.
- Ratio theories
- Logical relation theories
- Relative frequency theories
- Evaluating interpretations of probability

The phenomenon

- Generics and frequency statements are very common, and very mysterious.
- Weak, since they ar econtingent, and allow exceptions:

  (9) a. Every bird flies.
      b. Birds (usually) fly.

- Strong, since they are *lawlike*:

  (10) a. Every Supreme Court judge has a prime Social Security number.
       b. Supreme Court judges (always/usually) have a prime Social Security number.

A probabilistic account

(11) a. Birds always fly

b. $P(\mathbf{fly}|\mathbf{bird}) = 1$

(12) a. Birds never fly

b. $P(\mathbf{fly}|\mathbf{bird}) = 0$

(13) a. Birds sometimes fly

b. $P(\mathbf{fly}|\mathbf{bird}) > 0$

(14) a. Birds $\left\{\begin{array}{l}\text{usually}\\ \text{often}\\ \emptyset\end{array}\right\}$ fly

b. $P(\mathbf{fly}|\mathbf{bird}) >?$

Ratio Theories

- $P(A|B) = \dfrac{|A \cap B|}{|B|}$
- The probability for a fair coin to come up "heads": the ratio of observed tosses where the coin came up "heads."

Åquist *et al.* (1980)

Frequency adverbs are relations between sets.

(15) a. Birds always fly

b. $\mathbf{always}(\lambda x \mathbf{bird}(x), \lambda x \mathbf{fly}(x))$

Truth conditions

- $\mathbf{always}(A, B)$ is true iff $P(A|B) = 1$
- $\mathbf{very\text{-}often}(A, B)$ is true iff $P(A|B) \geq 0.9$
- $\mathbf{often}(A, B)$ is true iff $P(A|B) \geq 0.7$
- $\mathbf{fairly\text{-}often}(A, B)$ is true iff $P(A|B) > 0.5$
- $\mathbf{fairly\text{-}seldom}(A, B)$ is true iff $P(A|B) < 0.5$
- $\mathbf{seldom}(A, B)$ is true iff $P(A|B) \leq 0.3$
- $\mathbf{very\text{-}seldom}(A, B)$ is true iff $P(A|B) < 0.1$
- $\mathbf{never}(A, B)$ is true iff $P(A|B) = 0$
- $\mathbf{sometimes}(A, B)$ is true iff $P(A|B) \neq 0$
- Generics?

The meaning of probability

Ratio between sets

$$P(A|B) =_{\mathrm{def}} \frac{|A \cap B|}{|B|}.$$

Logical Relation Theories

- $P(A|B)$ is the ratio of possible worlds in which both A and B hold to those where B holds.
- The probability that a coin comes up "heads,": the ratio of worlds in which it comes up "heads" to worlds in which it is tossed.
- Of course, some measure function on worlds must be defined.

Probability Distribution over Worlds

Schubert and Pelletier (1989):

- *Birds fly* is true just in case in "most" pairs of possible worlds and birds, the bird flies in that world.
- "most" is to be interpreted in terms of some probability distribution favoring worlds w' similar to the actual world with regard to the "inherent" or "essential" nature of things.
- They leave open what this means.

Normality

Kratzer (1981) deals with qualitative probability judgments:

(16) a. There is a good possibility that Gauzner-Michl was the murderer.

b. There is, however, still a slight possibility that Kastenjakl was the murderer.

c. Gauzner-Michl is more likely to be the murderer than Kastenjakl.

d. It is probable that Gauzner-Michl was the murderer.

Modality

- Qualitative probability judgments are modal.
- Modality has two components:
  1. Modal base W: A set of accessible worlds, indicating the type of modality—logical, physical, epistemic, deontic, etc.
  2. Ordering source: if $w_1 \leq w_2$, then w_2 is closer to the ideal, or more "normal" than w_1.

Human necessity

- ϕ is humanly necessary iff it is true in all worlds closest to the ideal.
- More formally: for all $w \in W$ there is $w' \in W$ s.t.
  1. $w \leq w'$
  2. for all $w'' \in W$, if $w' \leq w''$ then ϕ is true in w''.
- *It is probable that Gauzner-Michl was the murderer* is true iff in all worlds in which events turned out in the normal, expected way, Gauzner-Michl was the murderer.

Human possibility

- ϕ is a human possibility iff $\neg\phi$ is not a human necessity.
- *There is a good possibility that Gauzner-Michl was the murderer* is true just in case it is not probable that Gauzner-Michl was not the murderer.

Slight possibility

- ϕ is slightly possible iff
  1. There is at least one $w \in W$ in which ϕ is true, and
  2. $\neg\phi$ is a human necessity.
- *There is a slight possibility that Kastenjakl was the murderer* means that it is probable that Kastenjakl was not the murderer, but it is still possible that he was.

Comparative possibility

- ϕ_1 is more possible than ϕ_2 iff for every world in which ϕ_2 holds, there is a world where ϕ_1 holds which is at least as normal, but there is a world where ϕ_1 holds for which there is no world at least as normal in which ϕ_2 holds.
- Formally:
  1. for all $w \in W$, if ϕ_2 holds in w then there is $w \leq w'$ s.t. ϕ_1 holds in w', and
  2. there is $w \in W$ s.t. ϕ_1 holds in w and for no $w \leq w'$, ϕ_2 holds in w'.
- The truth of *Gauzner-Michl is more likely to be the murderer than Kastenjakl* entails that for every world in which Kastenjakl is the murderer, there is a world at least as normal where Gauzner-Michl is; but there is at least one world w where Gauzner-Michl is the murderer, and in all the worlds that are at least as normal as w, Kastenjakl is not the murderer.

Generics

Many researchers:

- Generics express human necessity.
- For example: *Birds fly* is true just in case in the most normal worlds, all birds fly.
- Cannot account for frequency adverbs, because it cannot handle quantitative judgments.

Relative Frequency Theories

- $P(A|B)$ is the mathematical limit of the proportion of Bs that are As as the number of Bs approaches infinity.
- Intuitively: the longer you toss a coin, the closer to 0.5 will the ratio of "heads" get. If you could toss the coin an infinite number of times, you would get *exactly* 0.5.

Cohen (1999)

- Models with branching time.
- Each linear course of time is called a *history*.
- *Birds (usually) fly* is evaluated with respect to histories that admit infinite sequences of birds.
- *Birds (usually) fly* is true iff in every *admissible* such sequence, the limit of relative frequency of flying birds among birds is greater than 0.5.

Admissibility

- It is impossible to observe infinitely long sequences in the actual world.
- They must be extrapolated from the actual history.
- Hence, observed instances must provide a good statistical sample.
- But we don't know how the sample was selected, so any sufficiently large sample must be a good sample.
- Thus: admissible histories must contain a sufficiently long interval of the actual history, and the relative frequency over that interval must be close to that of the admissible history as a whole.

Examples

(17) a. Birds (usually) fly

b. In an admissible history: the proportion of flying birds remains roughly the same, forever.

(18) a. John (often) jogs in the park.

b. In an admissible history: John continues to jog with roughly the same frequency, forever.

Homogeneity

- After Salmon (1977): a reference class B is homogeneous iff for every partition of B, and every $B' \subseteq B$ induced by the partition, $P(A|B)$ is roughly equal to $P(A|B')$.
- If we consider only temporal partitions: The domain of generics and frequency adverbs is homogeneous.
- What about other partitions?

Generics vs. frequency adverbs

While frequency adverbs require their domain to be homogeneous only with respect to the time partition, generics require homogeneity with respect to a great number of other partitions as well. Some examples:

(19) a. LOCATION: Israelis (usually) live on the coastal plain.

b. AGE: People are (usually) over three years old.

c. SEX: Primary school teachers are (usually) female.

Evaluating Interpretations of Probability

- Which, if any, of the interpretations of probability should we choose?
- L. J. Cohen (1989): different interpretations of probability are appropriate for different applications.
- The applications are characterized by 4 parameters.

Necessity vs. Contingency

Generics and frequency statements are true or false contingently.

Ratio: Contingent.

Logical relation: Necessary; but if the actual world is favored—contingent.

Relative frequency: Contingent.

Propositions vs. Properties

- What are A and B in $P(A|B)$?
- Generics and frequency statements relate properties.
- And the theories?

  Ratio: Properties.

  Logical relation: Propositions. But can be ameneded to be about ratios of pairs of worlds and individuals, hence, in effect, properties.

  Relative frequency: Properties.

Substitutivity

- When can we substitute other terms for A or B without changing the probability $P(A|B)$? When they have the same extension? Or intension? Or...?
- Generics and frequency adverbs are not extensional:

  (20) A computer (always) computes the daily weather forecast (Carlson 1989).

  Suppose today's weather forecast predicts a severe blizzard, and is consequently the main news item. Yet the following is false:

  (21) A computer (always) computes the main news item.

Intensionality

- But they are not fully intensional either: Suppose that the weather report is John's favorite newspaper feature. Then the following would be true:

  (22) A computer (always) computes John's favorite newspaper feature.

- Similarly, the following have the same truth conditions:

  (23) a. The whale suckles its young.
  b. The largest animal on Earth suckles its young.

- Hence: Generics and frequency adverbs are parametric on time, but not on possible worlds; if two properties have the same extension throughout time, they can be freely exchanged in a generic or a frequency sentence *salva veritate*.

Theories of probability and substitutivity

Ratio: Fully extensional.

Logical relation: Fully intensional.

Relative frequency: Parametric on time (histories) but not possible worlds.

Extensibility

- Would $P(A|B)$ would remain the same if the number of Bs were greater than it actually is?
- Generics and frequency adverbs are extensible; "Birds (always) fly" would keep its truth value even if there were more birds than there actually are.
- And the theories?

  Ratio: Not extensible.

  Logical relation: Extensible.

  Relative frequency: Extensible.

Summary

	Ratio	Log. relation	Rel. frequency
Contingent	yes	yes*	yes
Relating properties	yes	yes*	yes
Parametric on time	no	no	yes
Extensible	no	yes	yes

Conditionals

- We have seen that generics and frequency adverbs are extensible: if there were more birds, the following would still be true.

  (24) Birds (usually) fly.

- So, if something that is not a bird *were* a bird, it would probably fly.
- So the following is true:

  (25) If Dumbo were a bird, he would probably fly.

- But what does it mean?

A probabilistic account?

- All sentences have a probability of being true.

  (26) P(It will rain tomorrow)

- We want the probability of the conditional to be conditional probability:

$$\text{P(if it rains tomorrow, the game will be canceled)} = \text{P(the game will be canceled} \mid \text{it rains tomorrow)}$$

But. . . surprise!

- Lewis (1976): such a definition of the semantics of the conditional is impossible in principle.
- Intuitive explanation:
- Suppose there were some proposition ϕ, representing the meaning of *If A then B* s.t. $P(\phi) = P(B|A)$.
- Then, there are two options:

1. $A \to B$ entails ϕ

- Then ϕ is true whenever A is false.
- Hence, ϕ must be probable when A is improbable.
- But now consider:

  (27) If it snows tomorrow, people will walk around in bathing suits.

  A is improbable, but still the conditional is also improbable.

2. $A \to B$ does not entail ϕ

- Then ϕ may be false when $A \to B$ is true.
- Hence, ϕ may be false when $A \wedge \neg B$ is false.
- But this is unintuitive.
- Example: suppose we have a bag containing black and red balls, and we draw two.
- $A=$ "The first ball is black."
- $B=$ "The second ball is red."
- Suppose we know that the following is **false**:

(28) a. Both balls are black.
b. $A \wedge \neg B$

- But then we must conclude:

(29) a. If the first ball is black, the second ball is red.
b. If A then B.

All is not lost

- Hence, we cannot associate with a conditional a proposition ϕ s.t. $P(B|A) = P(\phi)$.
- Does this mean we have to deny that conditionals have truth conditions at all (e.g. Adams 1975)?
- Not necessarily: $P(B|A)$ may still be equal, if not to the probability of the conditional, to some other value associated with it.

Stalnaker and Jeffrey (1994)

- All propositions are random variables.
- For non-conditionals, the random variable has only two values:
  - 1 for true,
  - 0 for false.
- For $A > B$:
  - If A is true and B is true, the random variable is 1
  - If A is true and B is false, the value of the random variable is 0.
  - If A is false, the value of the random variable is $P(B|A)$.

Expectation

- Since the value of a conditional is a random variable, it has an expected value:

  $E(A > B) =$
  $1 \times P(A \wedge B) + 0 \times P(A \wedge \neg B) + P(B|A) \times P(\neg A)$

- For many important cases, this expectation is equal to the conditional probability:

  $E(A > B) = P(B|A)$

Example

- $P(A) = 0.7$
- $P(A \wedge B) = 0.56$.
- $P(B|A) = 0.8$
- $P(A \wedge \neg B) = P(A) - P(A \wedge B) = 0.14$.
- The random variable $A > B$ is:
  - 1 if $A \wedge B$
  - 0 if $A \wedge \neg B$
  - 0.8 otherwise.
- So:

$$\begin{aligned} E(A > B) &= 1 \times 0.56 + 0 \times 0.14 + 0.8 \times 0.3 \\ &= 0.8 \\ &= P(B|A) \end{aligned}$$

Problematic example (Edgington 1991)

(30) a. If the match is wet, then if you strike it it will light.

b. $W > (S > L)$

- We want $E(W > (S > L)) = P(S > L|W)$.
- $P(S > L|W)$: given that the match is wet, the probability that it will light if struck.
- Intuitively, this is 0 (or close to it).
- Is the expected value also 0?

The Calculation

- Suppose the match is wet: W.
- Then if $S \wedge L$, $S > L$ is 1.
- Hence, the probability that $S > L$ is 1 is $P(S \wedge L|W)$.
- The probability that $S > L$ is 0 is $P(S \wedge \neg L|W)$
- The probability that $S > L$ is $P(L|S)$ is $P(\neg S|W)$. Therefore, the expected value is:

  (31) $1 \times P(S \wedge L|W) + 0 \times P(S \wedge \neg L|W) + P(L|S) \times P(\neg S|W)$

- Some numbers:
  - $P(W) = 0.55$.
  - $P(L|W) = 0$, therefore $P(S \wedge L|W) = 0$.
  - $P(L|S) = 0.9$
  - $P(S) = P(\neg S) = 0.5$
  - $P(W|\neg S) = 1$.
  - By Bayes's rule, $P(\neg S|W) = P(W|\neg S) \times \dfrac{P(\neg S)}{P(W)} = 1 \times \dfrac{0.5}{0.55} \approx 0.91$.
- So we get, instead of 0: $P(S > L|W) = 1 \times 0 + 0 + 0.9 \times 0.91 \approx 0.82$.

Vagueness

- It is hard to determine not only the truth conditions of conditionals, but even their truth $values$.
- Another phenomenon where judgments of truth value are hard is that of vague predication.
- For example, how tall is **tall**?

Kamp (1975)

- The goal: an account of comparatives in terms of positive adjectives.
- For example, $taller$ in terms of the meaning of $tall$.
- Requires a semantics where a predicate can hold of an entity to a certain degree.
- Then John is taller than Bill just in case the predicate $tall$ holds of John to a greater degree than it does of Bill.

Multi-valued logic?

- Proposal: represent degress as truth values between 0 and 1.
- But: how would the truth values be calculated compositionally?
- in some cases it's easy: $[\![\neg\phi]\!] = 1 - [\![\phi]\!]$.

Hard cases

What about $[\![\phi \wedge \psi]\!]$?

- Suppose that $[\![\phi]\!] = [\![\psi]\!] = 0.5$.
- Look for options:

1. $[\![\phi \wedge \psi]\!] = 0.5$?
   But then $[\![\phi \wedge \neg\phi]\!] = 0.5$, instead of 0.
2. $[\![\phi \wedge \psi]\!] = 0$?
   But then $[\![\phi \wedge \phi]\!] = 0$
3. No other option will work either

Partial models

- Classical models: the interpretation function F assigns to each symbol of arity n an n-place relation on the universe U.
- Three-valued logic: true, false, and undefined.
- Partial models: F assigns to each symbol of arity n an ordered pair of n place relations:

$$F(Q^n) = \langle F^+(Q^n), F^-(Q^n)\rangle.$$

- – $F^+(Q^n)$: Q^n definitely holds,
  – $F^-(Q^n)$: Q^n definitely does not hold
  – the rest: Q^n is undefined.
- For example:
  – $U = \{\text{Aleksandra}, \text{Bart}, \text{Caroline}\}$.
  – Aleksandra is definitely tall,
  – Caroline is definitely not tall,
  – Bart is neither definitely tall nor definitely not tall.
- $F(\mathbf{tall}) = \langle\{\text{Aleksandra}\}, \{\text{Caroline}\}\rangle$.

Completions

- Suppose M is a partial model and M' is a classical model.
- If M' agrees with M on F^+ and F^-, then M' is a $completion$ of M.
- For example:
  1. $F(\mathbf{tall}) = \{\text{Aleksandra}\}$.
  2. $F(\mathbf{tall}) = \{\text{Aleksandra}, \text{Bart}\}$.
- But not:

$$F(\mathbf{tall}) = \{\text{Aleksandra}, \text{Bart}, \text{Caroline}\}.$$

Supervaluation

- Let ϕ be a formula, and M a partial model.
- What is the truth value of ϕ?
  1. true, if ϕ is true in all completions of M
  2. false, if ϕ is false in all completions of M
  3. undefined, otherwise.

Tautologies and contradictions

- Suppose p is undefined in model M.
- In some completions, p is true, hence $\neg p$ is false.
- In other completions, p is false, hence $\neg p$ is true.
- So: in all completions, $p \vee \neg p$ is true.
- Hence, $p \vee \neg p$ is true in M.
- Similarly, $p \wedge \neg p$ is false in M.

Degrees of truth

- Associate with every sentence ϕ in model M the set of all completions that make it true.
- The set of all completions, if ϕ is true in M (or is a tautology)
- The empty set, if ϕ is false in M (or is a contradiction)
- Otherwise, intuitively: the larger the set associated with a sentence, the "truer" it is.
- To make this intuition precise, a probability function needs to be defined over these sets.

Vague models

A vague model $\mathcal{M}$ for a language L is a quadruple $\langle M, \mathcal{L}, \mathcal{F}, P\rangle$ where:

1. M is a partial model for L;
2. $\mathcal{L}$ is a set of completions of M;
3. $\mathcal{F}$ is a field of subsets over $\mathcal{L}$, s.t.
   for each formula $\phi \in L$ and assignment g,

   $$\{M' \in \mathcal{L} : [\![\phi]\!]^{M',g} = 1\} \in \mathcal{F}$$

4. P is a probability function over $\mathcal{F}$.

Truth degree

The degree to which ϕ is true is the value of the probability function over those completions where it is true:

(32) $[\![\phi]\!]^{\mathcal{M},g} = P(\{M' \in \mathcal{L} : [\![\phi]\!]^{M',g} = 1\}).$

Advantage #1 of using probability

- Suppose ϕ is definitely true in M, or is a tautology.
- Then it is true in all completions, $\mathcal{L}$.
- Since P is a probability function, $P(\mathcal{L}) = 1$.
- So, ϕ is definitely true in $\mathcal{M}$.

- Suppose ϕ is definitely false in M, or is a contradiction.
- Then it is true in no completion: $\emptyset$.
- Since P is a probability function, $P(\emptyset) = 0$.
- So, ϕ is definitely false in $\mathcal{M}$.

Recall that multi-valued logics cannot do this.

Advantage #2 of using probability

- We can decide arbitrarily whether p_1 is tall or not.
- But the decision whether p_2 is tall may no longer be arbitrary.
- Specifically, if p_1 is tall, and (33) is true, then p_2 must be tall too.

  (33) $\mathbf{p}_2$ is taller than $\mathbf{p}_1$.

- In other words: there is no completion where p_1 is tall and p_2 is not tall.

  $$\{M' \in \mathcal{L} : [\![\mathbf{tall}(\mathbf{p}_1)]\!]^{M',g} = 1\} \subseteq$$
  $$\{M' \in \mathcal{L} : [\![\mathbf{tall}(\mathbf{p}_2)]\!]^{M',g} = 1\}$$

- Because P is a probability function, if $A \subseteq B$ then $P(A) \leq P(B)$.
- Hence: $[\![\mathbf{tall}(\mathbf{p}_1)]\!]^{\mathcal{M},g} \leq [\![\mathbf{tall}(\mathbf{p}_2)]\!]^{\mathcal{M},g}$.
- It is "truer" that $\mathbf{p}_2$ is tall than that $\mathbf{p}_1$ is tall.
- These are the truth conditions of (33).

Compositionality

- There is no general formula to calculate $P(\phi \wedge \psi)$ based on $P(\phi)$ and $P(\psi)$.
- Hence, there is no general way to calculate the truth value of $\phi \wedge \psi$ based on the truth values of ϕ and ψ.
- But if ϕ and ψ are independent, then

$$P(\phi \wedge \psi) = P(\phi) \times \ P(\psi).$$

- So, if ϕ and ψ are independent, then

$$[\![\phi \wedge \psi]\!] = [\![\phi]\!] \times [\![\psi]\!].$$

- But what does it mean for two propositions to be independent?
- This depends on your interpretation of probability.

Independence, proposal #1

- Vagueness is uncertainty: we are uncertain whether (34) is true or not.

  (34) John is tall.

- The probability of ϕ: our degree of belief that ϕ is true.
- ϕ is independent of ψ iff knowing ϕ gives us no rational reason to change our belief about ψ.
- For example, if we know that John is tall, this does not affect our belief about how intelligent he is; hence, (34) and (35) are independent of each other.

  (35) John is intelligent.

Is vagueness a form of uncertainty?

Edgington (1996):

- There is a difference between being uncertain about whether ϕ holds, to ϕ's holding to a certain degree.
- Suppose we would like to drink coffee, and have two options:
  1. Shirley will serve us either coffee or tea, with a probability of 0.5.
  2. Roger will definitely serve a drink which is a mixture of coffee and tea; it is indeterminate whether it should be called "coffee" or "tea".
- Clearly, the two choices are not equivalent.
- Having a beverage that is coffee to the degree 0.5 is not the same as having a 0.5 chance of drinking coffee.

Independence, proposal #2

Edgington (1996):

- $P(\phi|\psi)$ is the truth value of ϕ if ψ is definitely true.
- ϕ and ψ are independent iff $P(\phi) = P(\phi|\psi)$
- For example, suppose John is good at math to the degree d_1, and intelligent to the degree d_2.
- Then, if we decide that John is definitely good at math, this may affect the truth value of his being intelligent.
- Hence, the two statements are not independent.
- But if we decide that John is definitely tall, this would not affect our judgment of how intelligent he is.
- Hence, being tall and being intelligent are independent.

Intuitions

- Does Kamp's account of vagueness fit our intuition?
- Suppose that John is tall to the degree 0.5, and intelligent to the degree 0.5.
- If the properties of being tall and being intelligent are independent of each other, John has the property of being both tall and intelligent to the degree

  $$0.5 \times 0.5 = 0.25$$

- Is this reasonable?

Many

- *Many* is vague: how many is many?
- But it is also ambiguous.

  (36) Many Nobel laureates watched the Olympic games.

  1. Cardinal reading: the number of Nobel laureates who watched the Olympic games was large, compared with some norm n (false).
  2. The proportion of Nobel laureates who watched the Olympic games was large, compared with some norm k (true?)

- Formally:
  1. Cardinal: $\mathbf{many}(A, B)$ iff $|A \cap B| > n$
  2. Proportional: $\mathbf{many}(A, B)$ iff $\dfrac{|A \cap B|}{|A|} > k$

Symmetry

- The cardinal reading is symmetric: (36) entails (37).
- The proportional reading is not: (36) does not entail (37).

(37) Many people who watched the Olympic games were Nobel laureates.

The vagueness of *many*

- Where do the parameters n and k come from?
- Fernando and Kamp (1996): these values depend on our expectations.
- If we don't expect Nobel laureates to be interested in the Olympic games, small numbers or percentages will count as *many*.
- If we expect sports fans to be very interested in the Olympic games, much higher norms would be required to make (38) true.

(38) Many sports fans watched the Olympic games.

Expectation and probability

- "Many As are Bs" is true just in case it could well have been the case that fewer As are Bs.
- Something is expected iff its probability is high.
- What interpretation of probability is appropriate?
- The crucial issue: intensionality.

Is $many$ extensional or intensional?

- According to the definitions above, $many$ is extensional.
- But is it? Keenan and Stavi (1986):

(39) a. Many lawyers attended the medical association meeting last year.

b. Many doctors attended the medical association meeting last year.

Even if all the lawyers were doctors and all the doctors were lawyers the truth values of the two sentences might differ.

- Kamp and Reyle (1993):

(40) a. Many houses in X burned down last year.

b. Many houses in X were insured against fire last year.

Solution

- Once the values of n and k are fixed, $many$ is extensional.
- But these same values depend on the intensions of the arguments of $many$.

Logical relation theory

- We need a fully intensional theory of probability.
- Ratio theory is fully extensional—out.
- Relative frequency theory is only partly intensional—out.
- Logical relation theory is in.

The cardinal reading

- There is some number n s.t. there are at least n individuals that are both ϕ and ψ, and this n counts as many:
- $\mathbf{many}_x(\phi(x), \psi(x))$ iff $|\phi(x) \wedge \psi(x)|_{x,w_0} \geq n \wedge n\text{-}\mathbf{is\text{-}many}_x(\phi(x), \psi(x)))$.
- Where $|\alpha(x)|_{x,w}$ indicates the number of individuals that satisfy α in world w:

$$|\{u : [\![\alpha(x)]\!]^{w,g[u/x]} = 1\}|.$$

Probability and $\mathbf{is\text{-}many}$

- n-$\mathbf{is\text{-}many}_x(\phi(x), \psi(x))$ iff

$$P(\{w : |\phi(x) \wedge \psi(x)|_{x,w} < n\}) > c$$

- In words: take the probability of the set of worlds where the number of individuals satisfying both ϕ and ψ is less than n; this probability is greater than some parameter c.
- Hence: $\mathbf{many}_x(\phi, \psi)$ means that there is some number n, s.t. there are n individuals that are both ϕ ans ψ, and there could well have been fewer than n.
- This reading is a function of (the intension of) $\phi \wedge \psi$, hence symmetric.

The proportional reading

- The only difference is in the definition of n-**is-many**.
- Conditional instead of unconditional probability:
- $n\textbf{-is-many}_x(\phi(x), \psi(x))$ iff

  $$P(\{w : |\phi(x) \wedge \psi(x)|_{x,w} < n\} \mid \{w : |\phi(x)|_{x,w} = |\phi(x)|_{x,w_0}\}) > c.$$

- In words: n is many iff there could well have been fewer than n xs that satisfy $\phi(x) \wedge \psi(x)$, given that there are $|\phi(x)|_{x,w_0}$ xs that satisfy $\phi(x)$.
- ϕ has a privileged position (the reference class), conditional probability), hence this reading is asymmetric.

Back to the k parameter

$\mathbf{many}_x(\phi(x), \psi(x))$ is true iff

$$\frac{|[\![\phi]\!] \cap [\![\psi]\!]|}{|[\![\phi]\!]|} > k,$$

where

$$k = \frac{n}{|\phi|} \wedge n\text{-}\mathbf{is\text{-}many}.$$

Focus and the proportional reading

(41) a. Many linguists arrived [by bus]$_F$ (de Hoop and Solà 1995).

b. $\neq$A high proportion of the total population of linguists arrived by bus.

c. $=$ A high proportion of the linguistis who arrived, arried by bus.

Focus Semantic Value

Rooth (1985): Every expression ϕ has

- an ordinary semantic value $[\![\phi]\!]^O$;
- a focus semantic value $[\![\phi]\!]^F$, representing alternatives to the focus.

$$[\![\text{[John]}_B \text{ loves [Mary]}_F]\!]^F = \{\text{John loves Mary}, \\ \text{John loves Kate}, \\ \ldots\}$$

Solution

- Geiluß (1993) and de Hoop and Solà (1995): The union of the focus semantic value restricts the quantifier.
- $\mathbf{many}_x(\phi(x), \psi(x))$ is true iff

$$\frac{|[\![\phi]\!]^O \cap [\![\psi]\!]^O|}{|[\![\phi]\!]^O \cap \cup [\![\psi]\!]^F|} > k$$

- Suppose

$$[\![\text{arrived [by bus]}_F]\!]^F = \{\text{arrived by bus}, \\ \text{arrived by car}, \\ \text{arived by train}, \ldots\}.$$

- Then $\cup [\![\text{arrived [by bus]}_F]\!]^F = \text{arrived}$.
- $\mathbf{many}_x(\mathbf{linguist}(x), \mathbf{arrrived\text{-}by\text{-}bus}(x))$ is true iff

$$\frac{|\mathbf{linguist} \cap \mathbf{arrived\text{-}by\text{-}bus}|}{|\mathbf{linguist} \cap \mathbf{arrived}|} > k$$

Additional reading?

Cannot account for:

(42) Many Scandinavians have won the Nobel prize in literature.

Proposed solutions:

1. The Reverse Interpretation view (Westerståhl 1985):

   (43) =Many winners of the Nobel prize in literature were Scandinavians.

2. The cardinality interpretation (de Hoop and Solà 1995): context determines what is a large number.

Problem #1: dependence on the size of the domain

The number of Scandinavians is predicted not to matter, but it does:

(44) Many Andorrans have won the Nobel prize in literature.

Problem #2: conservativity

- A generalized quantifier Q is conservative iff $Q(A, B) \Leftrightarrow Q(A, A \cap B)$.

(45) a. Most/all/some/no alligators like to sunbathe.

b. =Most/all/some/no alligators are alligators that like to sunbathe.

- The cardinality view: $many$ is conservative. But:

(46) Many Scandinavians are Scandinavians who have won the Nobel Prize in literature.

- The Reverse Interpretation view: $many$ is conservative with respect to its second argument. But:

(47) Many Scandinavians who have won the Nobel Prize in literature, have won the Nobel Prize in literature.

Relative readings

- The number of Scandinavians matters, so this is a proportional reading.
- The reading is non-conservative, so non-Scandinavians matter too.
- Relative proportional reading:

(48) a. Many Scandinavians have won the Nobel Prize in literature.

b. =The proportion of Scandinavians who have won the Nobel Prize in literature is greater than the probability that an arbitrary person has won a Nobel Prize in literature.

Intonation

The sentence appears most felicitous in a contrastive context, with contrastive accent:

(49) A: Have many Israelis won the Nobel Prize in literature?
B: No, many SCANDINAVIANS have won the Nobel Prize in literature.

Is this a contrastive focus (Herburger 1997) or contrastive topic?

Focus or topic?

- Strawson (1964): topics presuppose their descriptive content.

  (50) a. *The King of France is bald.
  b. The exhibition was visited by the King of France.

  This also applies to contrastive topics:

  (51) A: The President of France is bald.
  *B: No, the KING of France is bald.

  Now consider:

  (52) a. *Many MARTIANS have won the Nobel Prize in literature.
  b. Many Scandinavians have won the Nobel Prize in silly walks.

- de Hoop and Solà (1995):

  (53) a. Russia has the greatest number of scientists in the world, but...
  b. ...few of the people in Russia are scientists.
  c. ...*few SCIENTISTS are in Russia.

- Hence: SCANDINAVIANS is a contrastic topic.
- It is pronounced with a B-accent (fall-rise).

B Semantic value

Cohen (to appear): alternatives to the B-accented element.

$$[\![[\text{John}]_B \text{ loves } [\text{Mary}]_F]\!]^B = \{\text{John loves Mary}, \text{Bill loves Mary}, \ldots\}$$

Contrast Semantic Value

Büring (1997; 1999): alternatives to both focus and the B-accented element.

$$[\![[\text{John}]_B \text{ loves } [\text{Mary}]_F]\!]^{B+F} = \{ \{\text{John loves Mary}, \text{John loves Kate}, \ldots\}, \{\text{Bill loves Mary}, \text{Bill loves Kate}, \ldots\}, \ldots \}$$

Definition of relative reading

- $\dfrac{|[\![\phi]\!]^O \cap [\![\psi]\!]^O|}{|[\![\phi]\!]^O \cap \cup [\![\psi]\!]^F|} > k'$
- What is k'?
- Proposal: take the left-hand side, and replace every expression with the union of its B semantic value:

$$\frac{|[\![\phi]\!]^O \cap [\![\psi]\!]^O|}{|[\![\phi]\!]^O \cap \cup [\![\psi]\!]^F|} > \frac{|\cup [\![\phi]\!]^B \cap \cup [\![\psi]\!]^B|}{|\cup [\![\phi]\!]^B \cap \cup\cup [\![\psi]\!]^{B+F}|}.$$

- Now turn the right-hand side from a proportion to a probability:

$$\frac{|[\![\phi]\!]^O \cap [\![\psi]\!]^O|}{|[\![\phi]\!]^O \cap \cup [\![\psi]\!]^F|} > P(\cup [\![\psi]\!]^B | \cup [\![\phi]\!]^B \cap \cup\cup [\![\psi]\!]^{B+F})$$

- Note: $\cup [\![\psi]\!]^B \subseteq \cup\cup [\![\psi]\!]^{B+F}$

Example

- $\phi =$ [Scandinavians]$_B$
- $\psi =$ [won the Nobel Prize in literature]$_F$.
- $[\![\phi]\!]^O = \textbf{Scandinavian}$
- $[\![\phi]\!]^B = \{\textbf{Scandinavian}, \textbf{Briton}, \textbf{Israeli}, \ldots\}$
- $\cup [\![\phi]\!]^B = \textbf{person}$
- $[\![\psi]\!]^O = \cup [\![\psi]\!]^B = \textbf{win-lit-Nobel}$
- $[\![\psi]\!]^F = \{\textbf{win-lit-Nobel}, \textbf{win-Olympic-medal}, \textbf{win-Eurovision}, \ldots\}$
- $\cup [\![\psi]\!]^F = \cup\cup [\![\psi]\!]^{B+F} = \textbf{win-award}$

$$\frac{|\textbf{Scandinavian} \cap \textbf{win-lit-Nobel}|}{|\textbf{Scandinavian} \cap \textbf{win-award}|} > P(\textbf{win-lit-Nobel}|\textbf{person} \cap \textbf{win-award})$$

Unifying relative and absolute readings

- The value of the parameters k and k' appear very different and unrelated.
- We would like to find a unifying source for both.

Extension semantic value

- $[\![\phi]\!]^E$: a set of extensions of ϕ, one for each world.
- If ϕ is a property, $[\![\phi]\!]^E$ is a set of sets of individuals.
- Assume Lewis's (1968, 1971, 1986) counterpart theory, so that the individuals in different worlds are different.
- $\cup [\![\phi]\!]^E$: the set of all individuals that are in the extension of ϕ at some world.

Combining the extension and focus semantic value

- $[\![\phi]\!]^{E+F}$ is a set of sets; each set is the focus semantic value of ϕ at some world.
- If ϕ is a property, $\cup\cup[\![\phi]\!]^{E+F}$ is a set of individuals; each one is a member of the extension of some focus alternative to ϕ at some world.

Back to expectation

- $P(\cup [\![\psi]\!]^E | \cup [\![\phi]\!]^E \cap \cup\cup [\![\psi]\!]^{E+F})$
- The probability that something is a ψ in some world, given that it is a ϕ in some world and an alternative to ψ in some world.
- Since individuals in different worlds are different, this is the probability that if an individual in some world is a ϕ and an alternative to ψ, then it is a ψ.
- This is precisely the expectation that a relevant ϕ is a ψ.
- For example:

  (54) Many Nobel laurates watched [the Olympic games]$_F$ on TV.

- The probability that someone who is a Nobel laurate and watched something on TV in some world, watched the Olympic games in that world.

Redefining the absolute proportional reading

- $\mathbf{many}_x(\phi(x), \psi(x))$ is true iff

$$\frac{|[\![\phi]\!]^O \cap [\![\psi]\!]^O|}{|[\![\phi]\!]^O \cap \cup [\![\psi]\!]^F|} > P(\cup [\![\psi]\!]^E | \cup [\![\phi]\!]^E \cap \cup\cup [\![\psi]\!]^{E+F})$$

- Note what happens if we change E to B:
- $\mathbf{many}_x(\phi(x), \psi(x))$ is true iff

$$\frac{|[\![\phi]\!]^O \cap [\![\psi]\!]^O|}{|[\![\phi]\!]^O \cap \cup [\![\psi]\!]^F|} > P(\cup [\![\psi]\!]^B | \cup [\![\phi]\!]^B \cap \cup\cup [\![\psi]\!]^{B+F})$$

- This is precisely the relative reading!

The origin of the relative reading

- Why does *many* have a relative reading as well as an absolute reading?
- Because it is almost completely vague; in the right context, any proportion wil qualify as many (Partee 1988).
- Hence, hearers must apply strategies to "precisify" it.
- There are two such strategies:
  1. Looking at alternative worlds—absolute reading
  2. Looking at alternative values for the B-marked element—relative reading.

Back to generics

Generics also have relative readings:

(55) a. [Dutchmen]$_B$ are [good]$_F$ sailors.

b. $\neq$The probability for a Dutch sailor to be good is greater than 0.5 (absolute reading).

c. $=$The probability for a Dutch sailor to be good is higher than the probability for a sailor of arbitrary nationality to be good.

Why 0.5?

- Why does the absolute readng of generics require the probability to be greater than 0.5?
- In the case of $many$ we assume a probability measure, which favors worlds that are expected over worlds that are unexpected.
- But generics are not parametric on possible worlds, hence the probability does not favor one world over another.
- The expected value is then simply the mathematical expectation of a random number between 0 and 1: 0.5.

Even

- Fernando and Kamp (1996) view expectation as probability.
- Can we give a probabilistic acount of explicit expressions of expectation?
- For example:

(56) a. The food was so good, that even Denise finished everything on her plate.

b. $\Rightarrow$ We do not expect Denise to finish everything on her plate.

Chierchia and McConnell-Ginet (1990)

(57) a. Even N S

b. $\exists x(x = \text{N'} \wedge \text{S'})$

The logical form of (56.a) is:

(58) $\exists x(x = \mathbf{d} \wedge \mathbf{finish\text{-}all\text{-}food}(x))$,

It is satisfied just in case

(59) $[\![\mathbf{finish\text{-}all\text{-}food}(x)]\!]^{M,g[\text{Denise}/x]} = 1$

The presupposition of *even*

- A salient probabiity function P
- A salient set of individuals A
- For every $a' \in A$, if $a' \neq a$ then:

$$P(\llbracket S' \rrbracket^{M,g[a'/x]} = 1) > P(\llbracket S' \rrbracket^{M,g[a/x]} = 1)$$

- Thus, (56.a) presupposes that everybody is more likely to finish all the food on their plate than Denise is.

Is probability appropriate for *even*?

Francescotti (1995):

- (60) a. Granny was accused of kidnapping, and even murder.
  b. Granny was accused of murder, and even kidnapping.
- Fact: murder is more common than kidnapping.
- Hence: the probability that Granny committed murder is higher than of her committing a kidnapping.
- Hence: (60.b) ought to be felicitous, and (60.a) ought to be odd.
- But the facts are exactly the reverse...

Answer #1: probability is irrelevant

Kay (1990):

- The relevant notion is significance, not probability.

  (61) A: It looks as if Mary is doing well at Consolidated Widget. George [the second vice president] likes her work.
  B: That's nothing. Even Bill [the president] likes her work.

- Fine even in a context where there is no reason to think that Bill is less likely than George to like Mary's work.
- Or is it?

Answer #2: pick the sample space with care

- If the sample space contains various criminal acts, then murder is more likely than kidnapping.
- But supposes:
  - We classify actions according to how much trouble they cause the performer.
  - Actions that cause the same amount of trouble are grouped into an equivalence class.
  - The sample space contains these equivalence classes.
  - Under the assumption that people are not likely to cause themselves more trouble than is necessary, the equivalence class containing murder is **less** likely than the one containing kidnapping.
- If the sample space is restricted to equivalence classes of Granny's actions, the probability will depend on her character: if she tends to get in a maximal amount of trouble, then, again, murder is more likely than kidnapping.

Indirect Use of Probability

- The previous lectures discussed several cases where probability is explicitly introduced into the semantics.
- We will now look at a semantic account that does not use probability as such, but is inspired by probabilistic notions.
- van der Does and van Lambalgen (2000): a logic of perception reports.

Veridicality

- Perception reports are not veridical.

  (62) I see this arm.

- The following is **not** its logical form:

  (63) $\mathbf{see}(\mathbf{I}, x)$

- Given an assignment function that assigns the individual arm in question to x, (63) is satisfied just in case I see this arm.
- But for (62) to be true, all that is required is that something would appear like this arm to me: this arm, or another arm, or a leg, or a loaf of bread, or a hallucination.
- If my vision is perfectly reliable: (63) follows from (62).
- If my vision is completely unreliable: all that follows is

  (64) $\exists x\, \mathbf{see}(\mathbf{I}, x)$

Marr's (1982) theory of vision

The same object may be be represented at various levels of detail:

- A blur where nothing is distinguished;
- a sort of cylindrical shape;
- two cylinders, corresponding to the forearm and the upper arm;
- three cylinders, corresponding to the upper arm, the forearm, and the hand;
- additional cylinders, corresponding to the fingers;
- ...
- a perfectly detailed picture of an arm.

Partial knowledge

- Perfect knowledge: the variable is free.

  $\mathbf{see}(\mathbf{I}, x)$

- No knowledge: the variable is bound.

  $\exists x \ \mathbf{see}(\mathbf{I}, x)$

- Suppose: at the current level of detail, I cannot distinguish an arm from a leg, yet I am able to distinguish an arm from a loaf of bread.

- We do know something about the thing I see, but not its exact identity.

- Partial knowlege: the variable is "partially" bound?

Conditional expectation

- Used when we have only partial knowledge of the value of a random variable.
- Let X be the height of a person picked at random.
- If our vision is infinitely accurate, we have perfect information of the value of X.
- If our vision is very poor, we have no knowledge, and our best guess is the expectation of X.
- What if we can distinguish whether the person is tall, of middle height, or short?
  - If the person is tall— the expected value of X among tall people.
  - If the person is of middle height— the expected value of X among middle-height people.
  - If the person is short— the expected value of X among short people.

The math

- Let $\mathcal{G}$ is an algebra generated by the sets of tall people, middle-height people, and short people.
- Our guess of the height of the person is the conditional expectation of X given $\mathcal{G}$:

$$E(X|\mathcal{G}).$$

- $E(X|\mathcal{G})$ is itself a random variable, with three possible values:
  1. the average height of tall people
  2. the average height of medium-height people
  3. the average height of short people.
- It is "smoother" than X: it filters out distinctions that are real, but cannot be perceived.

Conditional quantification

A counterpart of conditional expectation:

- Let M be a model.
- $\mathcal{F}$ the set of assignments on M.
- $\mathcal{G}$ an algebra of subsets of $\mathcal{F}$.
- ϕ a formula.
- Identify with a formula ϕ the set of assignments that make it true:

  $$\{f \in \mathcal{F} : [\![\phi]\!]^{M,f} = 1\}.$$

- Then: the conditional quantifier is

  $$\exists(\phi|\mathcal{G})$$

  and it corresponds to the set of assignments

  $$\cap\{\mathcal{C} \in \mathcal{G} | \phi \subseteq \mathcal{C}\}$$

The idea

- $\exists(\phi|\mathcal{G})$ is the best estimate of the assignments that make ϕ true on the basis of the information available in $\mathcal{G}$.
- $\mathcal{G}$ contains those propositions (sets of assignments) whose truth we can verify at the current level of detail.
- So, an assignment makes $\exists(\phi|\mathcal{G})$ true just in case it makes true those statements entailed by ϕ that we can verify at the current level of detail.

An example

(65) a. I see this arm.

b. $\exists(\mathbf{arm}(x)|\mathcal{G})$

The algebra $\mathcal{G}$ represents the (visual) knowledge the speaker has.

- Suppose there are three individuals: a, l, and b
- We know that a is an arm, l is a leg, and b is a loaf of bread.

1. Totally reliable vision

- $\mathcal{G}$ is the algebra of the power set of $\mathcal{F}$, the set of all assignments.
- So, an assignment will make (65.b) true just in case it makes true all its entailments that the speaker can verify.
- But since the speaker can verify everything, an assignment will make (65.b) true just in case it will make $\mathbf{arm}(x)$ true, as desired.
- $\bigcap\{\mathcal{C} \in \mathcal{G} | \mathbf{arm}(x) \subseteq \mathcal{C}\} =$
  $\bigcap\{\mathcal{C} \subseteq \mathcal{F} | \mathbf{arm}(x) \subseteq \mathcal{C}\} =$
  $\mathbf{arm}(x) =$
  $\{f : f(x) = a\}$.

2. Totally unreliable vision

- The algebra $\mathcal{G}$ will be simply $\{\emptyset, \mathcal{F}\}$.
- The set of assignments that make (65.b) true will be those that make true all the entailments of $\mathbf{arm}(x)$ that the speaker can verify.
- But since the speaker can verify nothing, this will be the set of all assignments.
- $\bigcap\{\mathcal{C} \in \mathcal{G} | \mathbf{arm}(x) \subseteq \mathcal{C}\} = \mathcal{F} =$ $\{f : f(x) = a$ or $f(x) = l$ or $f(x) = b\}$.
- So in a case of no information at all, we are not able to distinguish any object from any other object.

3. Partially reliable vision

- Suppose the speaker can identify the property $\mathbf{body\text{-}part}$, but not $\mathbf{arm}$.
- So the speaker is able to distinguish a from b, but not from l.
- The algebra will be generated by $\{\mathbf{body\text{-}part}(x)\}$.
- An assignment will satisfy (65.b) just in case it satisfies its entailments that the speaker can verify, namely $\{\mathbf{body\text{-}part}(x)\}$.
- $\cap\{\mathcal{C} \in \mathcal{G} | \mathbf{arm}(x) \subseteq \mathcal{C}\} = \{\mathbf{body\text{-}part}(x)\} = \{f(x) = a$ or $f(x) = l\}$.

The role of probability in semantics

- We have discussed the role that probability plays in the study of semantic phenomena.
- Probability appears as a tool that aids traditional truth conditional semantics, not as a replacement for it.
- But could we have a more radical use of probability, one that plays a role at the fundamentals of the theory?

Traditional semantics

- The meaning of a sentence is a function from situations (possible worlds) to truth values.
- The intuition: in order to demonstrate understanding of the meaning of a sentence, one must be able to judge its truth or falsity in any situation.
- But to judge a sentence with certainty, the description of the situation must be complete.
- We can't judge the truth of (66) reliably not because we don't understand it, but because we don't have completel information.

  (66) There is life on Mars.

Towards a more realistic model

- This is an idealized view.
- In practice, we demonstrate understanding of a sentence by judging its truth on the basis of incpmplete informaton.
- If we wish to make the definition of meaning more realistic, we would need to use probability at a fundamental level of our semantics.

Probability-based semantics?

- Judgments of probability, rather than judgments of truth values.
- Understanding of the meaning of a sentence is demonstrated by judging its likelihood in a given situation, not its truth.
- The meaning of a sentence is a function from states of knowledge (sets of possible worlds) to probabilities.
- Such a semantics has yet to be developed.
- Perhaps that's because we can't judge the probability that (67) is true

    (67) Probability-based semantics is better than truth conditional semantics.

References

Åquist, L., Hoepelman, J., and Rohrer, C. 1980. Adverbs of frequency. In *Time, Tense and Quantifiers*, ed. by C. Rohrer 1–18. Tübingen: Verlag.

Adams, E. 1975. *The Logic of Conditionals*. Dordrecht: Reidel.

Bosch, P., and van der Sandt, R. (eds.) 1999. *Focus: Linguistic, Cognitive, and Computational Perspectives*. Cambridge: Cambridge University Press.

Büring, D. 1997. *The Meaning of Topic and Focus—The 59th Street Bridge Accent*. London: Routledge.

—— 1999. Topic. In Bosch and van der Sandt 142–165.

Carlson, G. 1989. On the semantic composition of English generic sentences. In Chierchia *et al.* (1989) 167–192.

Chierchia, G., and McConnell-Ginet, S. 1990. *Meaning and Grammar*. Cambridge, Massachussetts: MIT Press.

Chierchia, G., Partee, B. H., and Turner, R. (eds.) 1989. *Properties, Types and Meaning*. Dordrecht: Kluwer.

Cohen, A. 1999. Generics, frequency adverbs and probability. *Linguistics and Philosophy* 22.221–253.

—— to appear. Fronted quantificational adverbs. In *Dislocated Elements in Discourse*, ed. by B. Shaer, W. Frey, and C. Mainborn. Berlin: Routledge.

Cohen, L. J. 1989. *An Introduction to the Philosophy of Induction and Probability*. Oxford: Clarendon Press.

van der Does, J. M., and van Lambalgen, M. 2000. A logic of vision. *Linguistics and Philosophy* 23.1–92.

Edgington, D. 1991. The mystery of the missing matter of fact. In *Proceedings of the Aristotelian Society, supplementary volume 65* 185–209.

—— 1996. Vagueness by degrees. In *Vagueness*, ed. by R. Keefe and P. Smith 294–316. Cambridge, Massachusetts: MIT Press.

Eells, E., and Skyrms, B. (eds.) 1994. *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge: Cambridge University Press.

Fernando, T., and Kamp, H. 1996. Expecting many. In *Proceedings of the 6th Conference on Semantics and Linguistic Theory*, ed. by T. Galloway and J. Spence 53–68 Ithaca, NY. Cornell University.

Francescotti, R. M. 1995. Even: The conventional implicature approach reconsidered. *Linguistics and Philosophy* 18.153–173.

Geiluß, J. 1993. Nominal quantifiers and association with focus. In *Proceedings of the Workshop on the Semantic and Syntactic Analysis of Focus*, ed. by P. Ackema and M. Schoorlemmer 33–41.

Herburger, E. 1997. Focus and weak noun phrases. *Natural Language Semantics* 5.53–78.

de Hoop, H., and Solà, J. 1995. Determiners, context sets, and focus. In *Proceedings of the 14th West Coast Conference on Formal Linguistics—WCCFL XIV* Stanford.

Kamp, H., and Reyle, U. 1993. *From Discourse to Logic*. Dordrecht: Kluwer Academic Publishers.

KEENAN, E. L., and STAVI, J. 1986. A semantic characterization of natural language determiners. *Linguistics and Philosophy* 9.253–326.

KRATZER, A. 1981. The notional category of modality. In *Words, Worlds, and Contexts: New Approaches to Word Semantics*, ed. by H. J. Eikmeyer and H. Rieser 38–74. Berlin: de Gruyter.

LEWIS, D. 1968. Counterpart theory and quantified modal logic. *Journal of Philosophy* 65.113–126.

—— 1971. Counterparts of persons and their bodies. *Journal of Philosophy* 68.203–211.

—— 1976. Probabilities of conditionals and conditional probabilities. *Philosophical Review* 85.297–315.

—— 1986. *On the Plurality of Worlds*. Oxford: Blackwell.

MARR, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman.

PARTEE, B. H. 1988. Many quantifiers. In *Proceedings of ESCOL* 383–402.

ROOTH, M. E. 1985. *Association with Focus*. University of Massachusetts at Amherst dissertation.

SALMON, W. C. 1977. Objectively homogeneous reference classes. *Synthese* 36.399–414.

SCHUBERT, L. K., and PELLETIER, F. J. 1989. Generically speaking, or using discourse representation theory to interpret generics. In Chierchia *et al.* (1989) 193–268.

STALNAKER, R., and JEFFREY, R. 1994. Conditionals as random variables. In Eells and Skyrms (1994).

STRAWSON, P. F. 1964. Identifying reference and truth values. *Theoria* 30.

WESTERSTÅHL, D. 1985. Logical constants in quantifier languages. *Linguistics and Philosophy* 8.387–413.