Alternative-Sensitivity of *Likely* and *Probable*: Linguistic and Psychological Implications

Daniel Lassiter

Dept. of Linguistics, New York University
and
Institute of Philosophy, University of London

ConSOLE 19
8 January 2011

A badly controlled experiment

- You've applied for a job where there are four other applicants, and you are all equally qualified. How would you rate the following as descriptions of your chances?
  - It is certain that you will get the job.
  - It is likely that you will get the job.
  - It is somewhat likely that you will get the job.
  - It is unlikely that you will get the job.

A badly controlled experiment

- You've applied for a job that you really want, but you just found out that someone else has been offered the position. You've been told confidentially that you'll get it if the other candidate withdraws, which happens (in the company's long experience) about one time in five. How do you rate the following:
  - It is certain that you will get the job.
  - It is likely that you will get the job.
  - It is somewhat likely that you will get the job.
  - It is unlikely that you will get the job.

Outline

- Experiments like this show that subjects' probability judgments are sensitive to alternatives.
  - This has been claimed to show that subjects are reasoning incorrectly about probability.
  - And taken as evidence that humans don't make use of rule-governed probabilistic inferences.
- I'll show that this conclusion relies on mistaken assumptions about the semantics of *likely* and *probable*.
- An independently motivated semantic analysis of these items predicts alternative-sensitivity.
- This suggests the possibility that other evidence against probabilistic reasoning may also be accounted for in semantic/pragmatic terms.

Classical Perspectives on Probability

- How do humans reason using uncertain information?

Pierre Laplace, *Essai Philosophique sur les Probabilities*, 1814

We see in this essay that the theory of probability is basically nothing but good sense reduced to calculation; it allows us to assess with precision that which clear minds feel by a sort of instinct, without often being able to recognize.

- Philosophers and psychologists mostly assumed that reasoning about uncertainty was probabilistic until the 1970's.

Modern Perspectives on Probability

- Kahneman & Tversky's work (1971 et seq.) casts doubt on this assumption. The upshot was a new consensus:

Slovic et al., "Cognitive processes and societal risk taking", 1976

It may be argued that we have not had the opportunity to evolve an intellect capable of dealing conceptually with uncertainty.

Steven Jay Gould, *Bully for Brontosaurus*

Our minds are not built (for whatever reason) to work by the rules of probability.

- Recent work in Bayesian cognitive science questions K&T's conclusion, but indirectly.

Standard Semantics for *Likely* and *Probable*

- The Usual Analysis:
  - *Probable* and *likely* are synonyms (e.g., Horn 1989, Kratzer 1991)
  - ϕ is likely/probable $\equiv$ ϕ is more likely/probable than $\neg\phi$ (Kratzer 1991)
  - $prob(\phi) > prob(\neg\phi) \equiv prob(\phi) > 0.5$ (everyone in psychology)

Alternative-Sensitivity

- The alternative-sensitivity of certain expressions of uncertainty was discovered by Teigen (1988) and Windschitl & Wells (1998) independently.
- Yalcin (2009) introduced the topic into the linguistics/philosophy literature and replicated some relevant experiments.

Experimental results: Yalcin (2009)

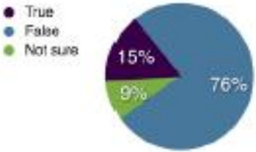
A soccer league has several teams competing against each other. Each team has a certain probability of winning, and of not winning, the championship. The chances of a certain team, Team X, are given below.

Team X	42 percent chance of winning the championship 58 percent chance of not winning the championship

Assuming the above information, is the following sentence true or false?

"Team X will probably win the championship."

Results as percentage of subjects:



Experimental results: Yalcin (2009)

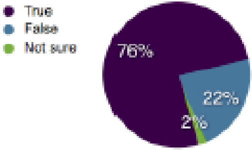
A soccer league has six teams competing against each other. Each team has a certain probability of winning the championship. The teams and their chances of winning are listed below.

Team A	12 percent chance of winning the championship
Team B	11 percent chance of winning the championship
Team C	13 percent chance of winning the championship
Team D	42 percent chance of winning the championship
Team E	12 percent chance of winning the championship
Team F	10 percent chance of winning the championship

Assuming the above information, is the following sentence true or false?

"Team D will probably win the championship."

Results as percentage of subjects:



Experimental Results

- I've replicated Yalcin's experiment and gotten 92.5% agreement in the second condition (37/40).

Alternative-Sensitivity, Effect 1

An event may be rated as more probable when it is presented in contrast to a number of outcomes with similar or lower probability than when it (or another event with the same probability) is presented in contrast to a single focal outcome with much higher probability. (Teigen, Windschitl & Wells, Yalcin)

Experimental Results: Teigen 1988

- Teigen 1988 asked subjects to indicate "probable" winners among 20 entries in the European Song Contest:

Ten days before the finals in the European Song Contest were to take place in Bergen (May 1986), 99 students in an introductory psychology course were given lists of the 20 nations participating in the contest and were asked to estimate of guess the chances for each participant to be elected winner. At that time, the Song Contest was the central current event in Bergen and the chances of individual participants were publicly and privately discussed. ...
Subjects in *Group 2* ($n = 35$) were asked for each participant whether they thought it was a *probable* or *not probable* winner. There was also a third response alternative, *neither probable nor not probable*, for those cases where neither expression was felt to be appropriate.
For *Group 3* ($n = 33$) the response alternatives were *improbable, not improbable*, and *neither*.

Experimental Results: Teigen 1988

Group 2			**Group 3**		
Expression	Mean	*SD*	Expression	Mean	*SD*
probable	7.8	3.0	not improbable	6.7	2.3
not probable	8.4	3.8	improbable	9.7	3.8
neither	3.8	3.3	neither	3.6	3.7

Table: Mean number of countries (of 20) judged to be probable and improbable winners of the 1986 European Song Contest in Teigen (1988).

- Most striking result: an average of 7.8 "probable" winners.

Experimental Results

Alternative-Sensitivity, Effect 2

Multiple mutually exclusive events may be judged "probable" or "likely" when (i) they are all roughly equiprobable, and (ii) no other event is substantially more likely. (Teigen)

Psychological Interpretations

- Standard assumptions predict that subjects who are reasoning correctly should (a) never judge more than one mutually exclusive event "probable", and (b) ignore the distribution of alternatives completely.
- Subjects consistently violate both of these predictions. One possibility: they are not reasoning correctly.
- Teigen interprets his results as showing that subjects routinely violate the laws of probability.

Psychological Interpretations

- W&W argue that these results are the effect of an associative, non-rule-based system for verbal probability judgments.
  - This system is supposed to be governed by a "comparison to the strongest" heuristic.
  - It's not at all clear whether this explains Effect 2, how it could work semantically, or what it would mean for knowledge of language in general ...
- Both of these interpretations fall in line with
  - The usual assumption that unexpected results in reasoning experiments indicate that subjects are making mistakes.
  - Kahneman & Tversky's claim that humans can't/don't reason probabilistically.

A Semantic Interpretation

- However, alternative-sensitivity is problematic only on the assumption that "probable" means "more probable than not".
- My main claim: experimental results are explained by the semantics of *probable* and *likely*.

Main Hypothesis

Likely and *probable* are semantically sensitive to alternatives: like other relative adjectives, they are evaluated by comparing their argument to a set of contextually salient alternatives. In the case of *likely* and *probable* the alternatives are often, but not always, provided by the denotation of the current Question Under Discussion (QUD, Roberts 1996).

- The only "error" that these experiments reveal is that we were wrong about what *likely* and *probable* mean.

Semantics of Relative Adjectives

- *Likely/probable* pattern with relative adjectives like *tall* on tests for adjective type, e.g. degree modifiers (Kennedy & McNally 2005):
  - Jeffrey is very/extremely/??completely/#slightly/#half tall.
  - It is very/extremely/??completely/#slightly/#half likely that it will rain.
- See Lassiter 2010 for lots more tests like this.

Comparison Classes

- Relative adjectives are sensitive to COMPARISON CLASSES.
- But *tall* does not just mean "taller than average for C":
  - If the average is $5'6''$, someone who is $5'6.2''$ will be "taller than average for C", but not 'tall' (Fara 2000). This suggests:

(3) $[\![tall]\!] = \lambda X_{<e,t>}.\lambda x_e.x$ is SIGNIFICANTLY taller than average/normal/expected for X

- Note that this presupposes $x \in X$: *Harold is tall for a jockey* is bad if Harold is not a jockey.
- As with *tall*, ϕ is not 'likely' if it is just barely more likely than $\neg\phi$ (Yalcin 2010). So maybe we have:

(4) $[\![likely]\!] = \lambda p_{<s,t>}.p$ is SIGNIFICANTLY more likely than $\neg p$

Comparison Classes

(4) $[\![likely]\!] = \lambda p_{<s,t>}.p$ is significantly more likely than $\neg p$

- We could make *likely* look even more like *tall* by adding a comparison class:

(4′) $[\![likely]\!] = \lambda P_{<st,t>}\lambda p_{<s,t>}.p$ is significantly more likely than average/normal/expected for P

- One argument for (4) over (4′) is that *likely* does not allow overt CCs:

(5) ?? It is likely that it will rain for a summer's day.

- But note the presupposition of the comparison class: (5) may be out simply because the proposition *it will rain* is not an instance of a summer's day.

Focus-Sensitivity

- Does the lack of overt CCs show that *likely* and *probable* do not have a comparison class argument? I'll argue no.
  - First I'll give a linguistic argument for a comparison class argument.
  - Then I'll show that the troubling experimental results are explained on this assumption as well.
- Imagine a lottery with a million tickets, in which one individual, Mr. Burns, is determined to win and buys 300,000. The rest are evenly distributed among the inhabitants of Springfield. Contrast (6) and (7):

(6) It is likely that [MR. BURNS will win the lottery].

(7) It is likely that [Mr. Burns will WIN THE LOTTERY].

Focus-Sensitivity

- Beaver & Clark (2003, 2008) show that focus-sensitive expressions fall into several types (not exhaustive):
  - Grammatically focus-sensitive operators like *only*, which must C-command the focus;
  - Expressions like *always* that have an implicit domain argument, which often display FREE ASSOCIATION WITH FOCUS (FAF).
- FAF is supposed to be **pragmatic** association with focus.
  - Roughly, focus makes salient a set of propositional alternatives, which preferentially fills the implicit argument of *always* if nothing else does.

Diagnostics for Focus Type

- Grammatically focus-sensitive operators must c-command the focus. This is not necessary for FAF.

8a. Frank is who I always give money to.
Can mean: "I don't give money to anyone but Frank."
8b. Frank is who I only give money to.
Cannot mean: "I don't give money to anyone but Frank."

- *Frank* can associate with *always* via pragmatic FAF in (8a), but not *only* in (8b) because c-command is needed. Similarly:

9a. We should thank the man who Mary always took to the movies.
9b. We should thank the man who Mary only took to the movies.

Diagnostics for Focus Type

- *Likely* seems to pattern with *always* on this test:

(10) Mr. Burns is who is likely to win the lottery.

(11) We should kiss up to Mr. Burns, who is likely to win the lottery.

- Remember: FAF occurs with expressions like *always* that have implicit domain arguments.
- In the case of *likely*, I suggest, this is the set of propositions with which the complement clause is compared for likelihood: its comparison class argument.

Alternatives & QUD

- As in Beaver & Clark's treatment of *always*, the implicit argument is filled by the set of focus alternatives, a partition of the common ground corresponding to the complete and relevant answers to the Question Under Discussion (QUD, cf. Roberts 1996).
- Focus-sensitivity of *likely* follows from the fact that the examples respond to different QUDs:

(12) **Alt**(6): *Who will win?* $\longrightarrow$ {*Mr. Burns will win, Bart will win, Millhouse will win, ...* }

(13) **Alt**(7): *What will Mr. Burns do?* $\longrightarrow$ {*Mr. Burns will win, Mr. Burns will not win*}

Alternatives & QUD

- Why is the standard analysis so initially plausible?
- I suggest that decontextualized examples tend to be interpreted with respect to a default QUD.
  - Unless context or focus supplies another value, QUD defaults to $?\psi$.
  - So **Alt**(ψ) defaults to $\{\psi, \neg\psi\}$.
  - The default option is to compare a proposition to its negation.

Result 1: Alternative Sensitivity

- Consider Yalcin's experiment. Subjects in Condition 1 saw a context which compares winning ($\mathit{prob}(\phi) = .42$) with not winning ($\mathit{prob}(\neg\phi) = .58$).
  - This naturally suggests QUD $?\phi$ and comparison class $\{\phi, \neg\phi\}$.
  - ϕ is clearly not more likely than average for this comparison class, so we predict the answer “false”.
  - Most subjects gave this response in Condition A.
- Subjects in Condition 2 saw a context with $\mathit{prob}(\phi) = .42$ compared to five less likely options.
  - ϕ is more likely than average for this comparison class, so we predict the answer “true”.
  - Most subjects gave this response in Condition B.

Result 2: Multiple "Probable" Outcomes

- Teigen found subjects rated multiple outcomes "probable" ($\mu = 8.2/20$).
- C is presumably $\{x \text{ wins} \mid x \text{ in the ESC}\}$
- On my approach, all "probable" options must be significantly more likely than average for C.
- Since probabilities must sum to 1, this means that many/most other options must be considered very unlikely winners.
- This is a new prediction which I'm currently testing. If it holds up it will be a strong argument for my theory.

A Problem Avoided

- Worry: doesn't all of this predict that ϕ and $\neg\phi$ can both be likely in the same context (Yalcin 2009)?
- No: QUD is a partition of CG, and $\mathit{prob}(\mathrm{CG}) = 1$, so

$$\sum_{\phi \in \mathrm{QUD}} \mathit{prob}(\phi) = \mathit{prob} \bigcup_{\phi \in \mathrm{QUD}} \phi = 1$$

- If QUD $= ?\phi$, either ϕ or $\neg\phi$ must have probability at least 50%.
- So the 'standard' can't be $< 50\%$ in a context where both ϕ and $\neg\phi$ are relevant.
- If we look for a context where the standard is $< 50\%$, ϕ and $\neg\phi$ will not both be in QUD.

Implications for Reasoning Experiments

- Sensitivity to the distribution of alternatives in probability judgments does not necessarily show that subjects are inconsistent or illogical.
- What it shows is that our assumptions about the meanings of probability expressions were mistaken:
  - Alternative-sensitivity is expected for relative adjectives like 'likely', 'probable', and their derivatives.
  - Similar arguments can be made for other likelihood expressions analyzed by Teigen and Windschitl & Wells, e.g. "good chances".

Implications for Reasoning Experiments

- Does this does show that the people are good at probabilistic reasoning?
- No: lots of other results remain showing that people make what appear to be mistakes.
- But, there is a real need to scrutinize these results for unjustified semantic assumptions.
- It may turn out that people are not quite so bad at probabilistic reasoning as we have thought.