RSA w/ politeness

KNOWLEDGE OF LANGUAGE

/lɪŋˈɡwɪstɪks/

$[\![\text{Joe}]\!] = \lambda e . \lambda w . \text{Joe}(e, w)$

GENERAL WORLD KNOWLEDGE
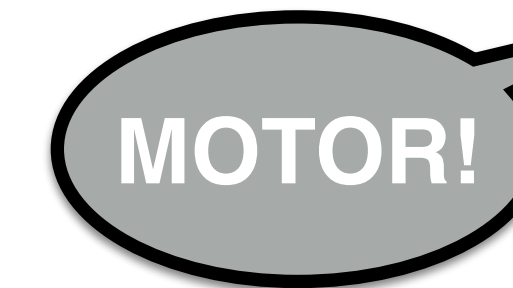
ACTUAL CONTEXT OF CONVERSATION

PRAGMATIC INTERPRETER

probability

ACTUAL SPEAKER

BIKE!

MODEL OF SPEAKER BEHAVIOR

MOTOR!

ALTERNATIVE UTTERANCES & INTERPRETATIONS
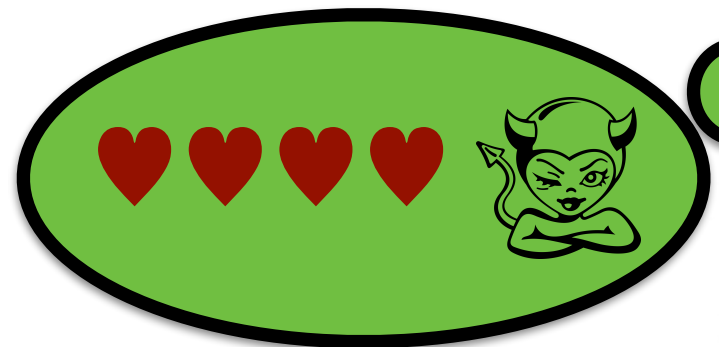
KNOWLEDGE OF LANGUAGE

/lɪŋˈɡwɪstɪks/

$[\![\text{Joe}]\!] = \lambda e \,.\, \lambda w \,.\, \text{Joe}(e, w)$
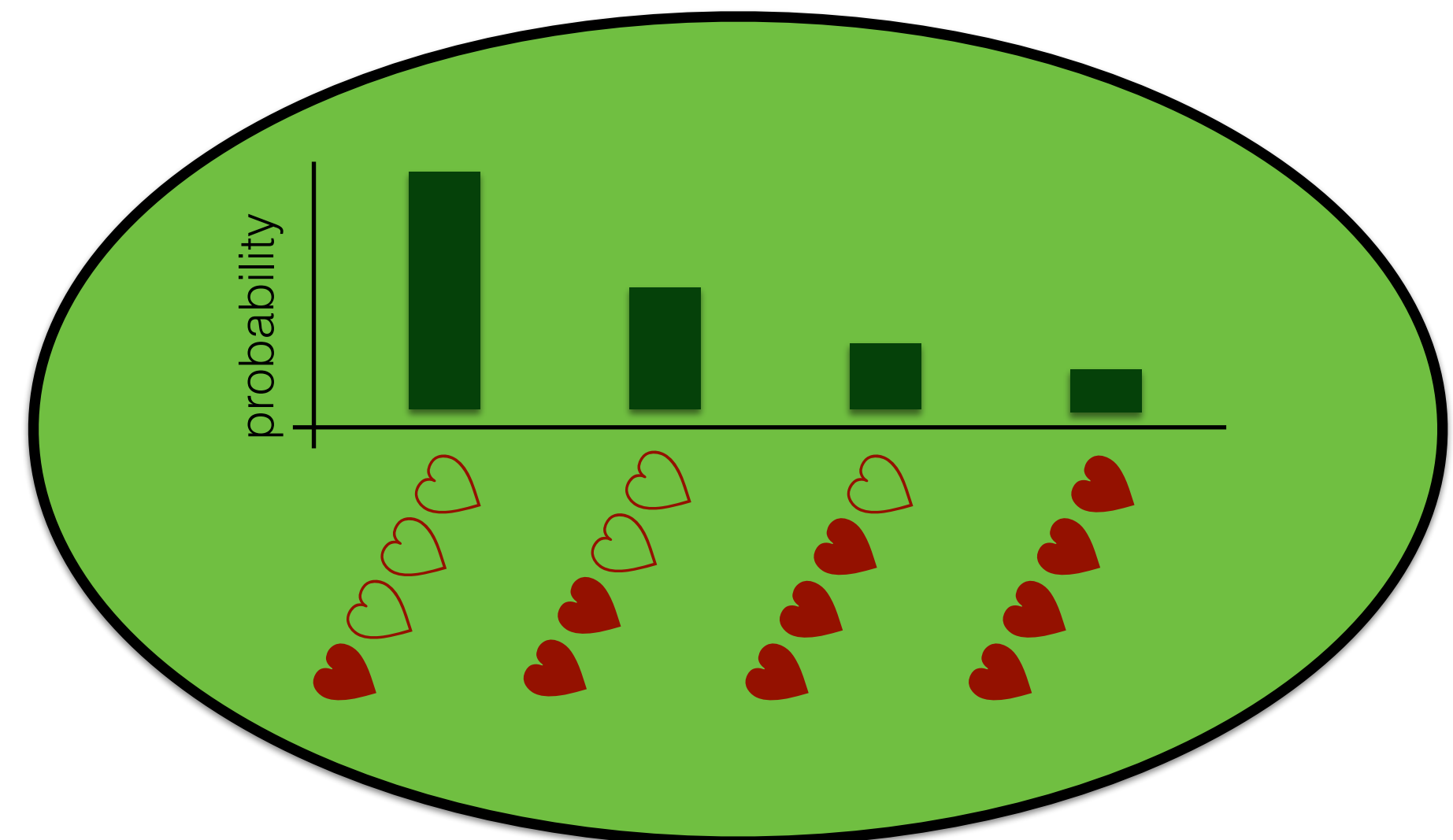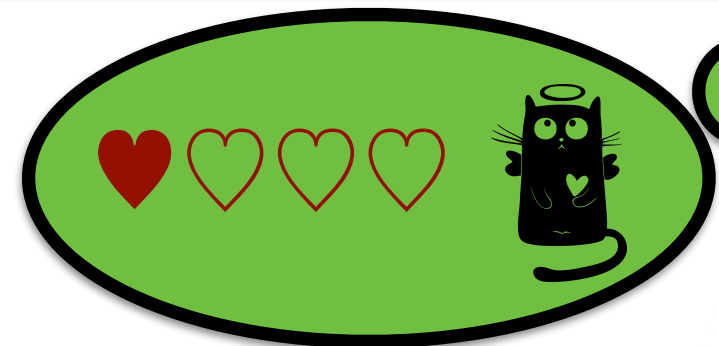
GENERAL WORLD KNOWLEDGE

ACTUAL CONTEXT OF CONVERSATION

probability

ACTUAL SPEAKER

"good"

MODEL OF SPEAKER BEHAVIOR

"great"
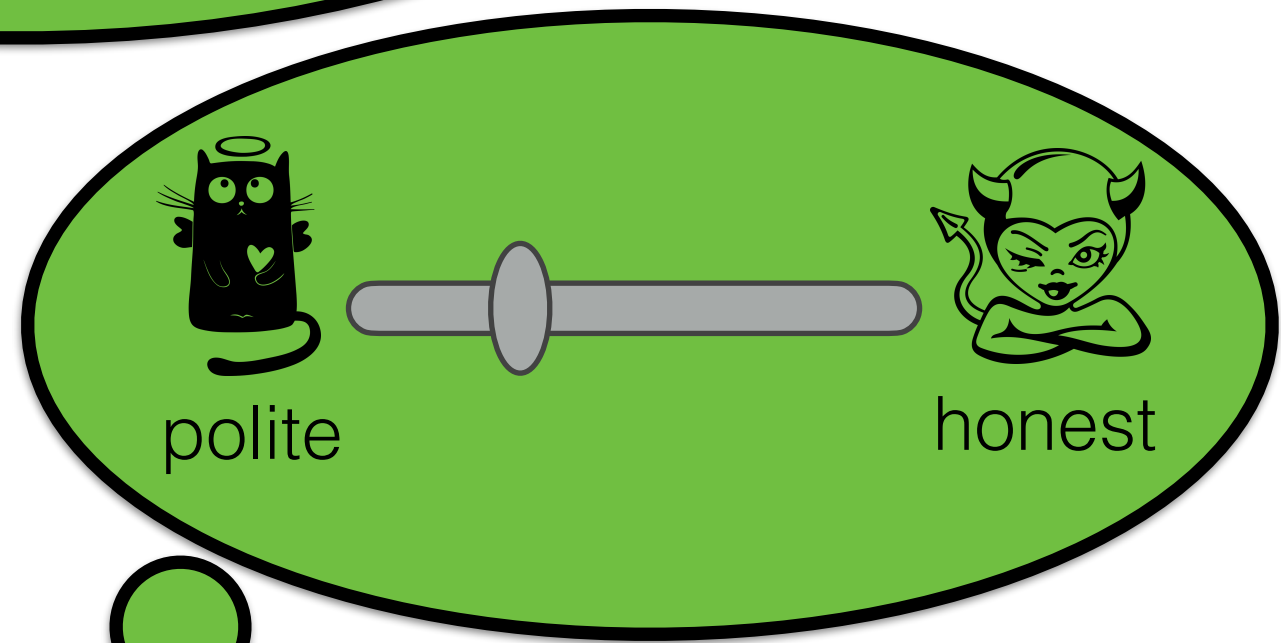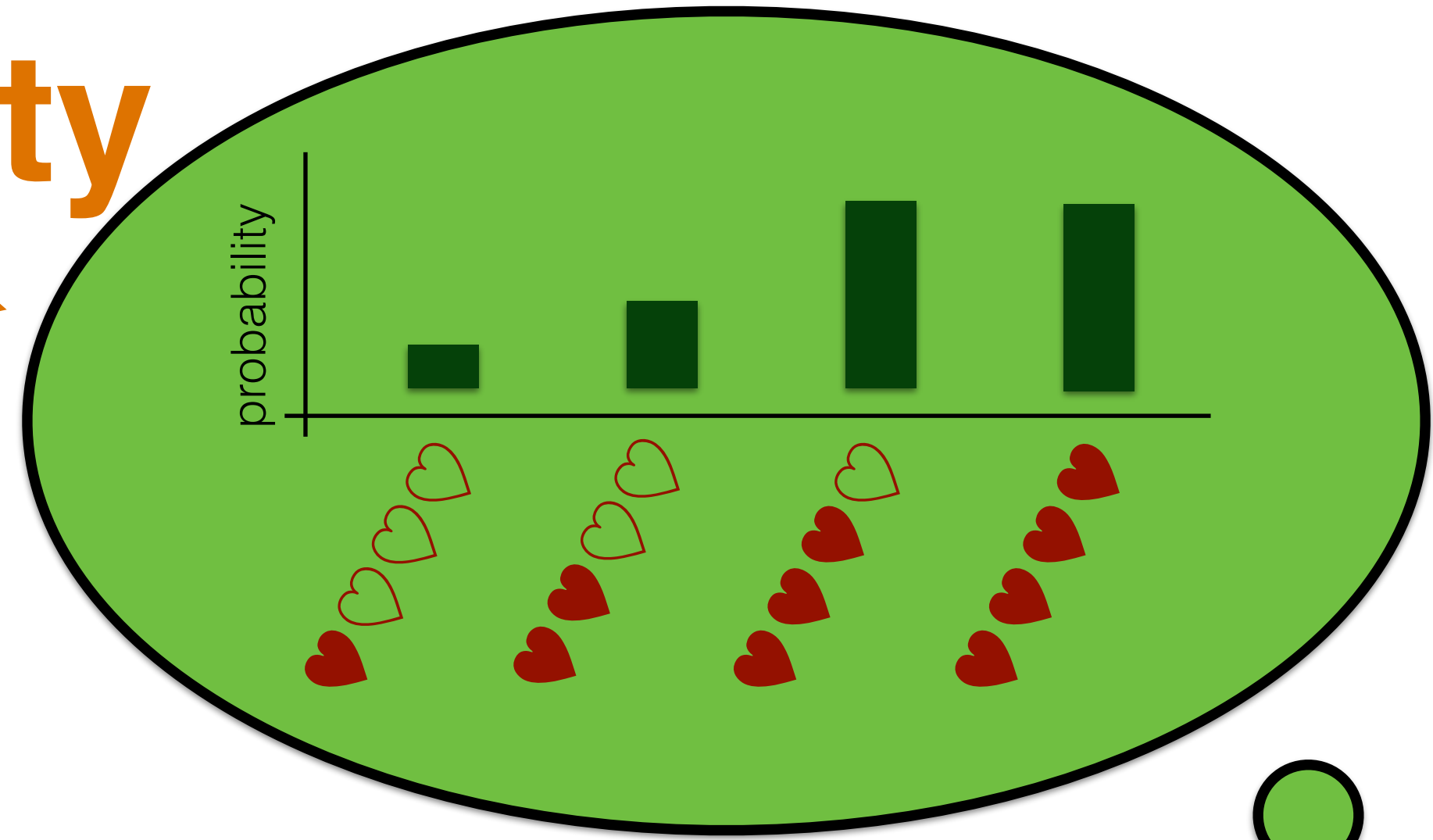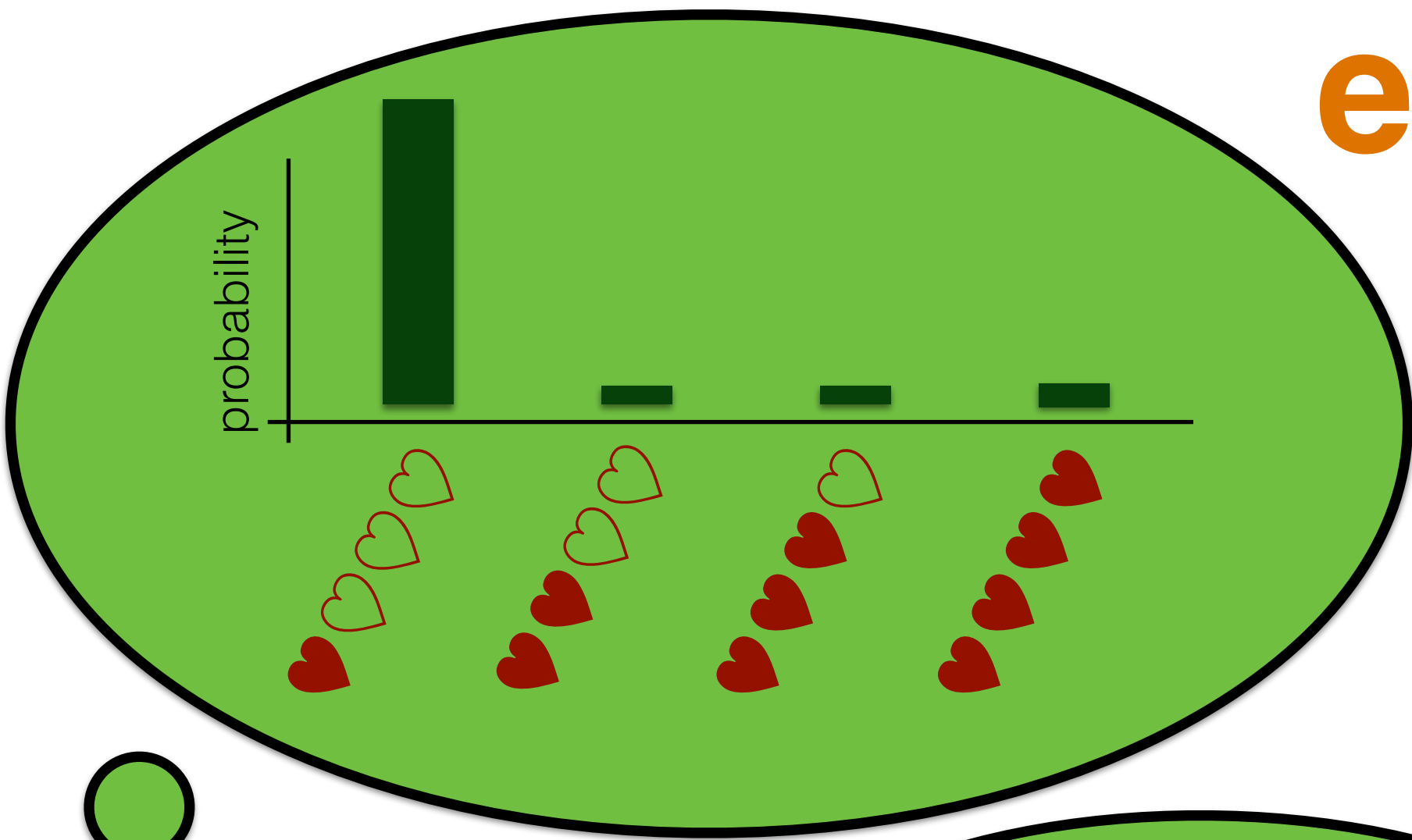
ALTERNATIVE UTTERANCES
& INTERPRETATIONS

PRAGMATIC INTERPRETER

epistemic utility

informativity
KL-divergence

probability

probability

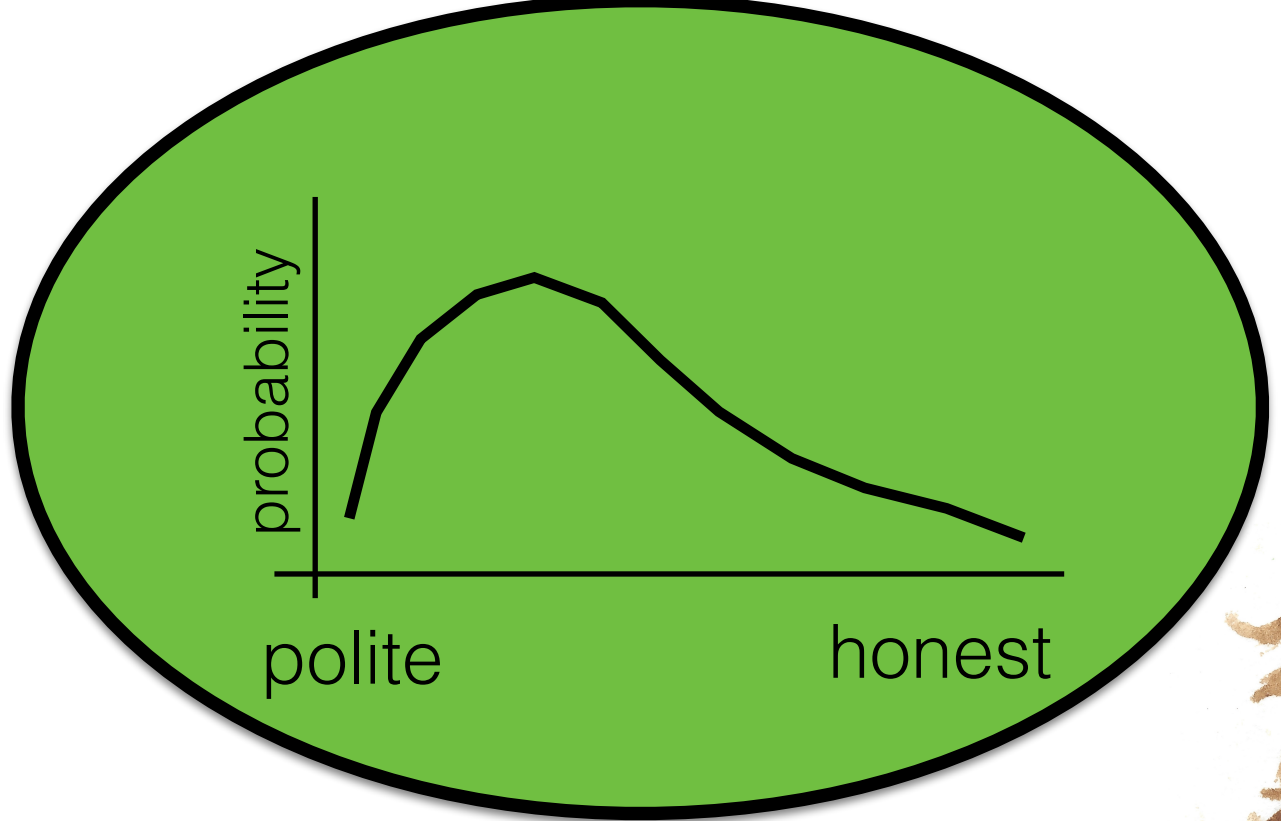linear
combination

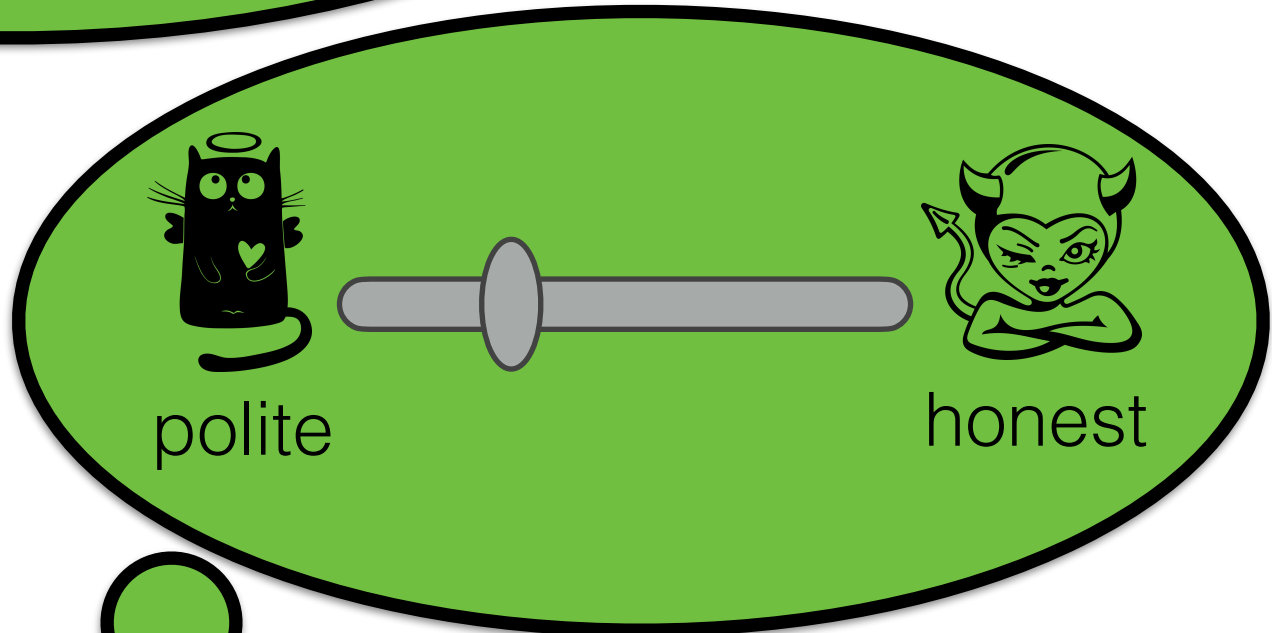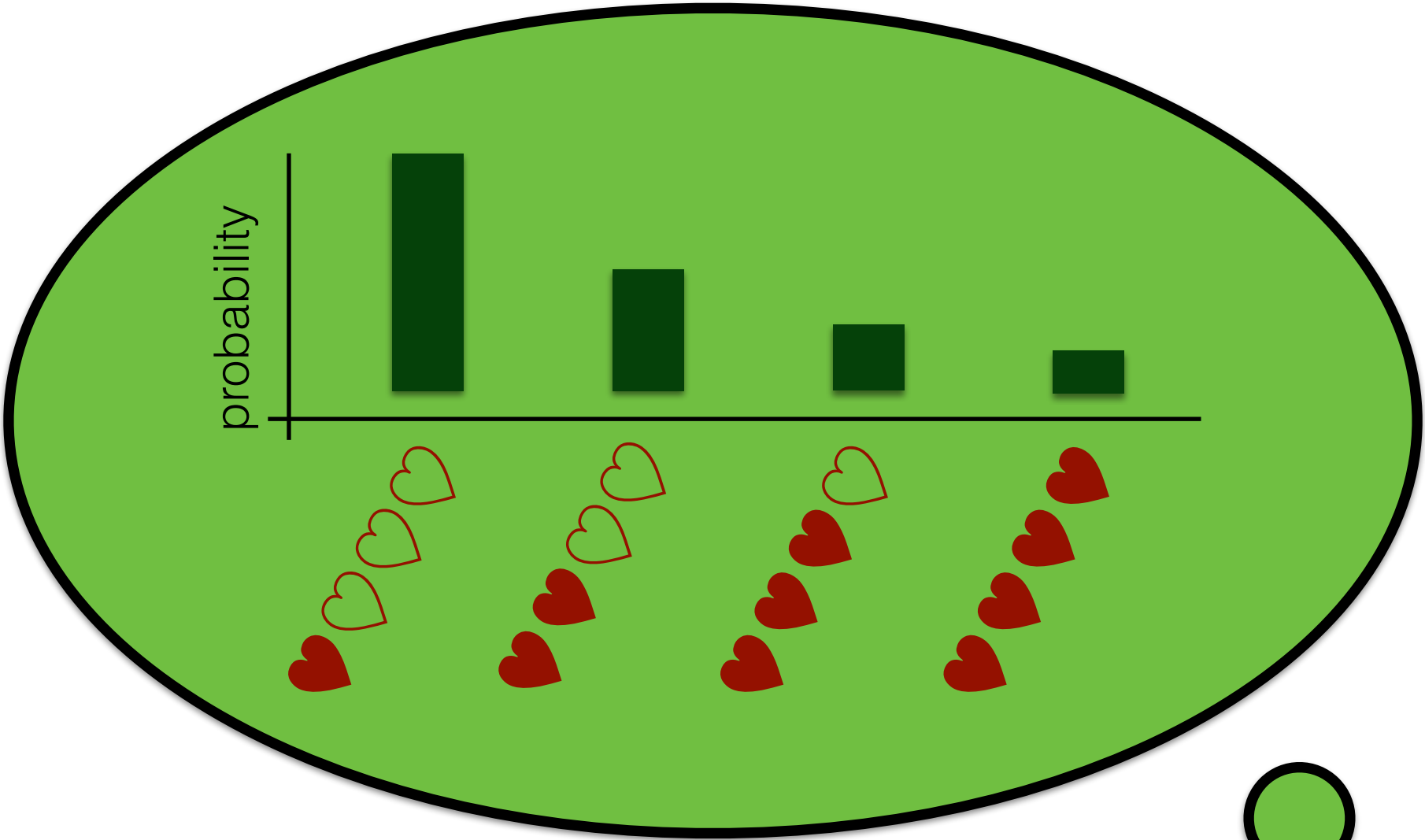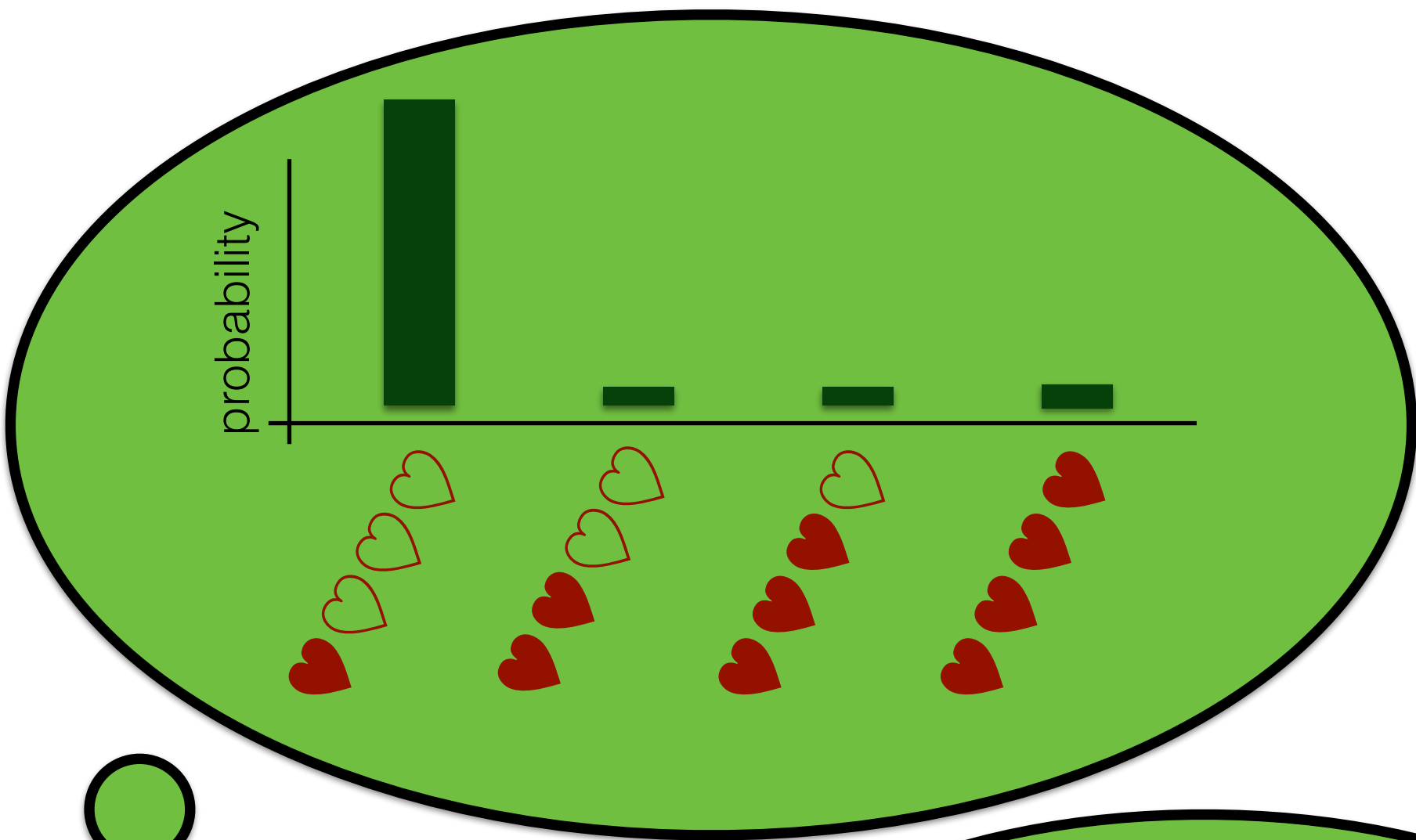polite   honest

social utility

PRAGMATIC SPEAKER

"good"

LITERAL INTERPRETER

PRAGMATIC SPEAKER

"good"

PRAGMATIC INTERPRETER

modeling polite language use

Vanilla RSA

$$P_{L_0}(s \mid u) = P(s \mid [\![u]\!])$$

$$P_{S_1}(u \mid s) \propto \exp\left(\alpha\left(\log P_{L_0}(s \mid u) - \mathrm{C}(u)\right)\right)$$

$$P_{L_1}(s \mid u) \propto P(s)\, P_{S_1}(u \mid s)$$

polite language model

$$P_{lit}(s \mid u) \propto L(u, s)$$

$$P_S(u \mid s, \varphi) = \text{SoftMax}\left(\alpha U(s, u, \varphi)\right)$$

$$U(s, u, \varphi) = \varphi \underbrace{\log P_{lit}(s \mid u)}_{\text{epistemic util}} + (1 - \varphi) \underbrace{\sum_{s'} P_{lit}(s' \mid u) \ V(s')}_{\text{social util}}$$

$$P_L(s, \varphi \mid u) \propto P(s) \ P(\varphi) \ P_{S_1}(u \mid s, \varphi)$$