Blurred Boundaries, Borderline Cases and the Utility of Vagueness

Jonathan Lawry

Department of Engineering Mathematics, University of Bristol, UK

SUM 2016

Talk Outline

Some people are always critical of vague statements. I tend rather to be critical of precise statements; they are the only ones which can correctly be labeled 'wrong'. (Raymond Smullyan)

- Vagueness and its utility.
- Lipman's result: The sub-optimality of vagueness.
- Multiple sender channels exploiting probabilistic vagueness.
- Can we escape Lipman?
- Mitigating the risks of making forecasts.
- Borderline cases and Hurwicz criterion.
- Vagueness as a route to building consensus.
- Conclusions.

What is Vagueness?

Today, vague predicates are standardly characterized by three main 'symptoms', namely as predicates that are sorities susceptible, that have borderline cases, and that have blurry boundaries. (Paul Egré)

- Vagueness is a multifaceted phenomenon making a single unified treatment difficult.
- We will focus on (versions of) two of the 'symptoms' identified by Egré.
- *Semantic Uncertainty (Blurred Boundaries):* Explicit representation of uncertainty about concept definitions as a consequence of the empirical manner in which language is learnt.
- *Indeterminism:* Borderline cases to which neither the concept or its negation absolutely apply.

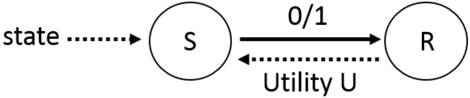
Is Vagueness Useful?

If you can't be kind, at least be vague. [Judith Martin]

- Vagueness pervades natural language yet it is frowned upon in the western scientific tradition.
- In science clarity and (semantic) precision are seen as being a fundamental prerequisite to progress.
- A hypothesis must be clearly formulated before it can be properly empirically tested.
- So why is vagueness so common in almost all aspects of language?
- Is it actually useful in communication and if so how and why?
- Can this study inform new applications to artificial intelligence?

Lipman: The Sub-optimality of Vagueness

it seems rather far fetched to conclude that we have simply tolerated a worldwide, several thousand year efficiency loss. [Barton Lipman]



- Pure Strategy Game: there are 2 strategies; 'transmit 0' and 'transmit 1'.
- Mixed Strategy Game: all probability distributions on $\{0, 1\}$.
- Pure strategies are special cases of mixed strategies: $0 : 1, 1 : 0$ and $0 : 0, 1 : 1$.
- Lipman's result tells that *for any state of the world, the maximal expected utility value is always obtain from a pure strategy.*

Probabilistic Approaches to Vagueness

- Probabilistic approaches to vagueness have a history dating back to work by Max Black, and include proposals by Loginov, Hisdal, Edgington, Borel and more recently Lawry, Lassiter and Egré.
- These models tend to be strongly related in that they assume an uncertain extension of a concept, where the uncertainty is quantified probabilistically.

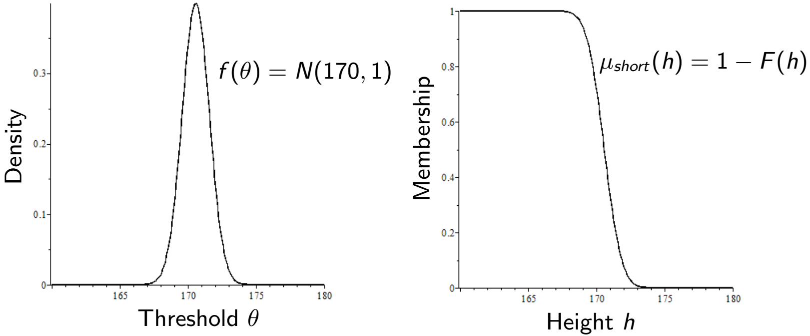
Example

Let $\Omega = \{Bill, Fred, Mary, Ethel\}$ be the domain of discourse. We then define a probability distribution of possible extensions of the concept happy as follows:

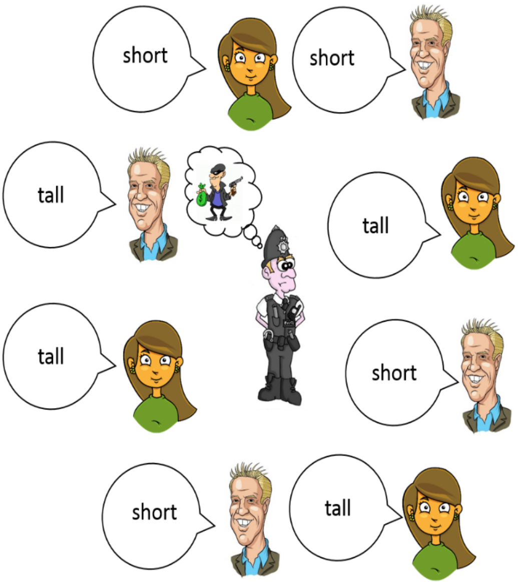
$$\{Bill, Ethel\} : 0.6,\ \{Bill, Ethel, Mary\} : 0.3,$$
$$\{Bill, Ethel, Mary, Fred\} : 0.1$$
$$\mu_{happy}(Mary) = 0.3 + 0.1 = 0.4$$

The Threshold Model

- Consider a gradable adjective defined on one dimensional scale like *short*.
- Suppose that the extension of *short* is the interval $[0, \theta]$ where θ is an uncertainty threshold.
- Suppose $\theta \sim f$ with cumulative distribution F.
- $\mu_{short}(h) = P(\{\theta : h \in [0, \theta]\}) = P(\theta \geq h) = 1 - F(h)$.

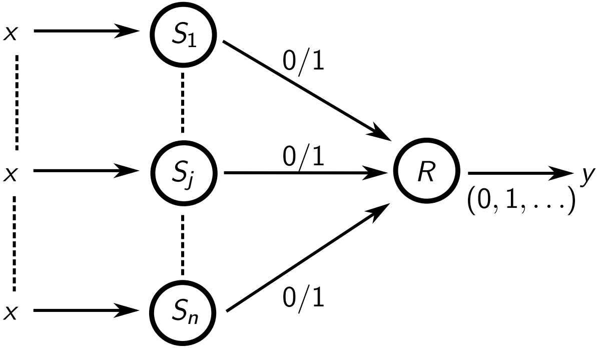


The Bristol Robbery

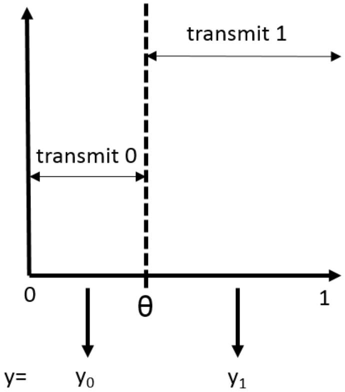


Multiple Sender Channels

- $x \sim U[0,1]$ described by two possible signals 0 or 1.
- Senders $S_1, \ldots, S_n$ each transmit a signal stochastically and the receiver R aggregates these to obtain an estimate y of x.
- Performance evaluated by expected squared error $\mathbb{E}((x-y)^2)$.



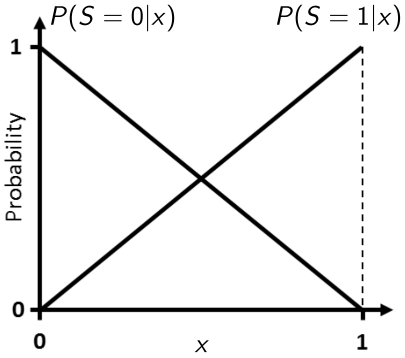
Boolean Channels



- *Optimal Boolean Channel:* $\mathbb{E}((x-y)^2)$ is minimal when $\theta = 0.5$, $y_0 = 0.25$ and $y_1 = 0.75$.
- In a boolean channel (without noise) all senders send the same signal.
- This is equivalent to a single sender channel.

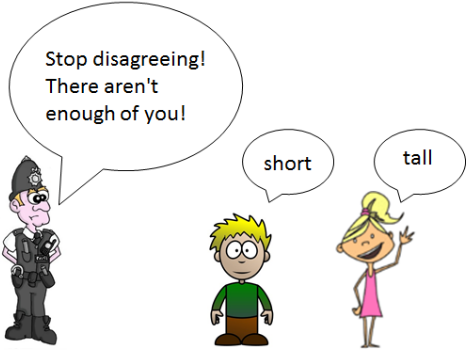
Stochastic Senders

- Suppose $\theta \sim f$ and take $P(S=0|x) = P(\theta \geq x) = 1 - F(x)$.
- Let $T = \sum_{i=1}^{n} S_i$ and $y = \mathbb{E}(x|y)$ i.e. the error minimizing estimator.
- Initially suppose that $\theta \sim U(0,1)$ so that $P(S=0|x) = 1 - x$ and $P(S=1|x) = x$.
- In this case $y = \frac{T+1}{n+2}$ i.e. Laplace's rule.



Comparing Vague and Boolean Channels

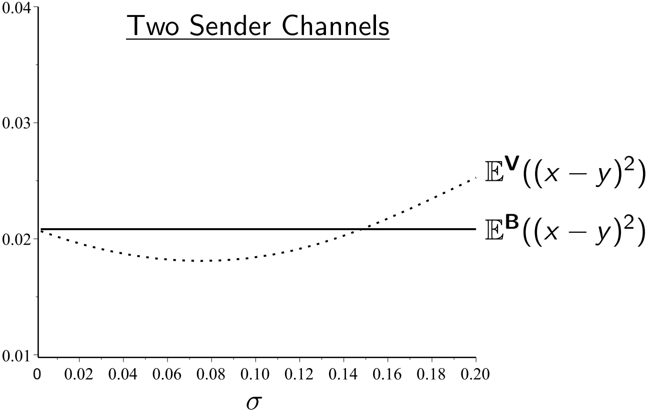
- For the vague channel $\mathbb{E}^{\mathbf{V}}((x-y)^2) = \frac{1}{6(n+2)}$.
- For the optimal boolean channel $\mathbb{E}^{\mathbf{B}}((x-y)^2) = \frac{1}{48}$.
- $\mathbb{E}^{\mathbf{V}}((x-y)^2) \leq \mathbb{E}^{\mathbf{B}}((x-y)^2)$ if and only if $n \geq 6$.
- But as yet this is not a particularly compelling case for vagueness.
- After all how often do we have the luxury of aggregating assertions from that many different independent sources?



Optimizing Membership Functions

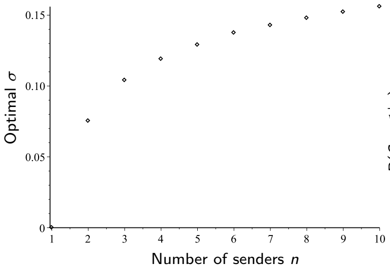
- Consider a more general parameterised family of memberships.
- Suppose θ is normally distributed with mean $\frac{1}{2}$ and s.d. σ.
- After normalisation this gives:

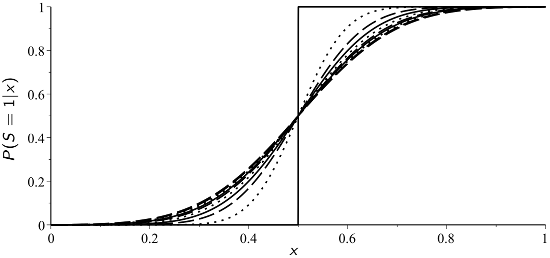
$$P(S=1|x) = \frac{1}{2}(1+\mathrm{erf}(\frac{x-\frac{1}{2}}{\sigma\sqrt{2}}) + (x-\frac{1}{2})(1+\mathrm{erf}(\frac{-\frac{1}{2}}{\sigma\sqrt{2}})).$$



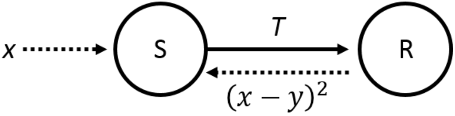
The More the Vaguer

- The value optimal value of σ increases with n but with decreasing differentials.
- Notice that we have distinctly sigmoid functions even for 10 senders or more.



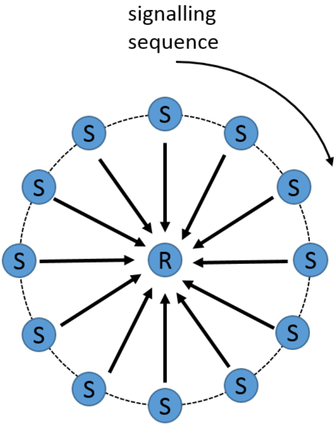


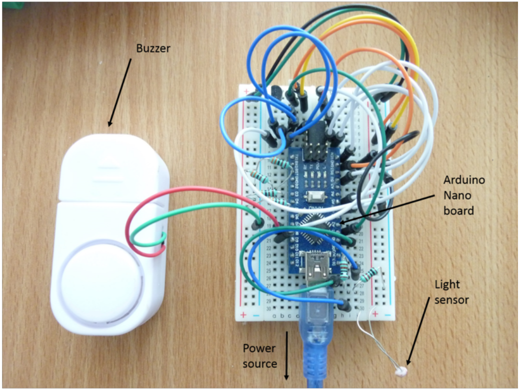
Escaping Lipman



- Think of S as an aggregate of $S_1, \ldots, S_n$.
- Pure Strategy Game: $T = 0, \ldots, n$.
- Mixed Strategy Game: All probability distributions on $\{0, \ldots, n\}$.
- Stochastic Game: All binomial distributions on $\{0, \ldots, n\}$.
- Notice that there is no stochastic strategy of the form $S = T$ for $T = 1, \ldots, n-1$.
- For $x \in (0,1)$ the optimal strategy is the pure strategy $T =$ the closest integer to nx.
- But this would require the senders to coordinate in order to send the best $n-$bit approximation of x.

Stochastic Sensor Networks





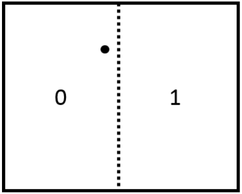
- Mean squared errors for normalized light levels:
- Stochastic 0.0191, Boolean 0.0223

Mitigating the Risk of Forecasts

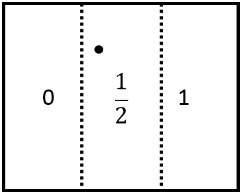
Earlier on today, apparently, a woman rung the BBC and said she heard there was a hurricane on the way... well, if you're watching, don't worry, there isn't!

- There can be advantages for using vague language in assertions and forecasts.
- For example, vague statement can be harder to falsify than crisp ones.
- More generally, we might expect...
- Vague assertions to be less likely to be false but to provide lower reward if true.
- Crisp assertions to be more likely to be false but to provide higher reward if true.

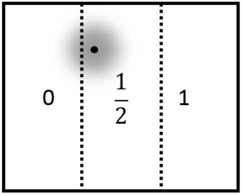
Borderline Cases and Uncertainty



Boolean



Three-valued



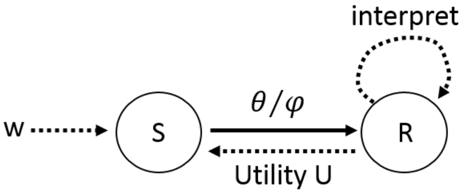
Three-valued + uncertainty

A Simple Propositional Model

- Let $\mathcal{L}$ be a language of propositional logic with propositional variables $p_1, \ldots, p_n$ and connectives $\neg$, $\wedge$ and $\vee$.
- Let $S\mathcal{L}$ denote the sentences of $\mathcal{L}$.
- *Three-valued Valuation*: A function $\mathbf{v} : S\mathcal{L} \to \{0, \frac{1}{2}, 1\}$ such that $\forall \theta, \varphi \in S\mathcal{L}$ if $\mathbf{v}(\theta) \in \{0, 1\}$ and $\mathbf{v}(\varphi) \in \{0, 1\}$ then $\mathbf{v}(\neg\theta) = 1 - \mathbf{v}(\theta)$, $\mathbf{v}(\theta \wedge \varphi) = \min(\mathbf{v}(\theta), \mathbf{v}(\varphi))$ and $\mathbf{v}(\theta \vee \varphi) = \max(\mathbf{v}(\theta), \mathbf{v}(\varphi))$.
- Let $\mathbb{V}$ be the finite set of all the possible three-valued valuations.
- Uncertainty is modelled by a probability distribution w on $\mathbb{V}$.
- *Lower and Upper Measures*: For $\theta \in S\mathcal{L}$, $\underline{\mu}(\theta) = w(\{\mathbf{v} \in \mathbb{V} : \mathbf{v}(\theta) = 1)$ and $\overline{\mu}(\theta) = w(\{\mathbf{v} \in \mathbb{V} : \mathbf{v}(\theta) \neq 0)$

Precisification and Vagueness

- A Sentence $\theta \in S\mathcal{L}$ is said to be *vague* if $\exists \mathbf{v} \in \mathbb{V}$ such that $\mathbf{v}(\theta) = \frac{1}{2}$.
- In contrast a sentence $\varphi \in S\mathcal{L}$ is said to be *crisp* if $\forall \mathbf{v} \in \mathbb{V}$, $\mathbf{v}(\varphi) \neq \frac{1}{2}$.
- For $\theta, \varphi \in S\mathcal{L}$, $\theta \preceq \varphi$ meaning φ is a precisification θ if and only if $\forall \mathbf{v} \in \mathbb{V}$, $\mathbf{v}(\theta) = 1 \Rightarrow \mathbf{v}(\varphi) = 1$ and $\mathbf{v}(\theta) = 0 \Rightarrow \mathbf{v}(\varphi) = 0$.



Better to be Vague? It depends what you mean?

Question

Given a probability distribution w on $\mathbb{V}$ and a crisp sentence φ, is there a vague sentence θ where $\theta \preceq \varphi$ and where $\mathbb{E}(U(\theta)) \geq \mathbb{E}(U(\varphi))$?

- There are two ways that the receiver can interpret a vague assertion θ:
  - I_1: θ is true i.e. $\mathbf{v}(\theta) = 1$.
  - I_2: θ is not false i.e. $\mathbf{v}(\theta) \neq 0$.
- The sender does not know which interpretation the receiver will adopt.
- Therefore, the sender cannot determine $\mathbb{E}(U(\theta))$ precisely.

Hurwicz Criterion

The Hurwicz criterion has been proposed as a decision rule in those situations where the utility obtained from certain actions in certain states of the world is not precisely known.

Definition

Let U^- be the minimum possible utility and U^+ be the maximum possible utility. Then the speaker should maximize:

$$H = \alpha U^- + (1-\alpha)U^+ \text{ where } \alpha \in [0,1]$$

In the case they are uncertain about the state of the world then they should maximize:

$$\mathbb{E}(H) = \alpha\mathbb{E}(U^-) + (1-\alpha)\mathbb{E}(U^+)$$

Here we take $U^-(\theta) = \min\{U(\theta|I_1), U(\theta|I_2)\}$ and $U^+(\theta) = \max\{U(\theta|I_1), U(\theta|I_2)\}$

Accuracy Based Utility

- Suppose that the utility received by the sender is dependent only on the accuracy of their assertion.
- The sender receives utility 1 if the receiver judges that their assertion holds, and 0 otherwise.
- This assessment depends on which interpretation the receiver applies:

$$U(\theta|I_1) = \begin{cases} 1 : \mathbf{v}(\theta) = 1 \\ 0 : \mathbf{v}(\theta) \neq 1 \end{cases} \quad \text{and} \quad U(\theta|I_2) = \begin{cases} 1 : \mathbf{v}(\theta) \neq 0 \\ 0 : \mathbf{v}(\theta) = 0 \end{cases}$$

- In this case: $U^{-}(\theta) = U(\theta|I_1)$ and $U^{+}(\theta) = U(\theta|I_2)$
- Let w_S be the probability distribution on $\mathbb{V}$ representing the speakers beliefs.
- In this case:

$$\mathbb{E}(H(\theta)) = \alpha \underline{\mu}_S(\theta) + (1-\alpha)\overline{\mu}_S(\theta)$$

More Accurate to be Vague?

- If φ is crisp $\underline{\mu}_S(\varphi) = \overline{\mu}_S(\varphi) = \mu_S(\varphi)$ and $\mathbb{E}(H(\varphi)) = \mu_S(\varphi)$.
- If $\theta \preceq \varphi$ then $\underline{\mu}_S(\theta) \leq \mu_S(\varphi) \leq \underline{\mu}_S(\varphi)$.
- The sender will transmit θ in preference to φ when $\mathbb{E}(H(\theta)) \geq \mathbb{E}(H(\varphi))$ if and only if

$$\alpha \leq \frac{\overline{\mu}_S(\theta) - \mu_S(\varphi)}{\overline{\mu}_S(\theta) - \underline{\mu}_S(\theta)}$$

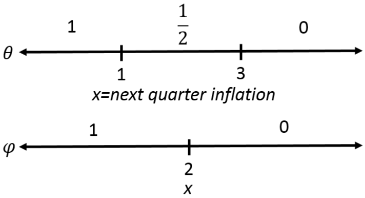
- This is an upper bound on α so:
- The more 'generous' or 'supportive' the receiver the more likely that it is better to transmit a vague than a crisp sentence.

Question

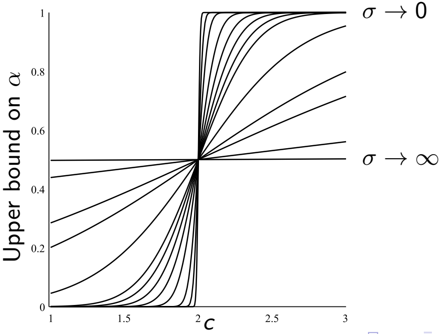
Is it better to assert a sentence about which the sender is completely uncertain, or one which they are certain is borderline? Suppose that $\mu_S(\varphi) = 0.5$ and $\underline{\mu}_S(\theta) = 0$ and $\overline{\mu}_S(\theta) = 1$ then θ is asserted if and only if $\alpha \leq 0.5$.

Example: One Dimensional Scale

Let $\theta =$'Next quarter inflation will be low' and $\varphi =$'Next quarter inflation will be at most 2%'



Suppose $x \sim N(c, \sigma)$



Utility Based on Accuracy and Informativeness

- In addition to accuracy the utility received by the sender should depend on how useful the information provided is to the receiver.
- A possible proxy for usefulness is informativeness i.e. how much information an assertion provides.
- Let $INF_1(\theta) \geq INF_2(\theta)$ quantify the information provided to the listener by assertion θ under interpretations I_1 and I_2 respectively.
- A possible definition for the sender's utility is then:

$$U(\theta|I_1) = \begin{cases} INF_1(\theta) : \mathbf{v}(\theta) = 1 \\ 0 : \mathbf{v}(\theta) \neq 1 \end{cases} \quad U(\theta|I_2) = \begin{cases} INF_2(\theta) : \mathbf{v}(\theta) \neq 0 \\ 0 : \mathbf{v}(\theta) = 0 \end{cases}$$

- From this we have that:

$$U^-(\theta) = \begin{cases} INF_2(\theta) : \mathbf{v}(\theta) = 1 \\ 0 : \mathbf{v}(\theta) \neq 1 \end{cases} \quad U^+(\theta) = \begin{cases} INF_1(\theta) : \mathbf{v}(\theta) = 1 \\ INF_2(\theta) : \mathbf{v}(\theta) = \frac{1}{2} \\ 0 : \mathbf{v}(\theta) = 0 \end{cases}$$

Is it more Informatively Accurate to be Vague?

- Taking into account informativeness then we obtain the following upper bound on α where θ is asserted in preference to φ:

$$\alpha \leq \frac{INF_2(\theta)\left(\overline{\mu}_S(\theta) - \underline{\mu}_S(\theta)\right) + INF_1(\theta)\underline{\mu}_S(\theta) - INF(\varphi)\mu_S(\varphi)}{INF_2(\theta)\left(\overline{\mu}_S(\theta) - \underline{\mu}_S(\theta)\right) + \left(INF_1(\theta) - INF_2(\theta)\right)\underline{\mu}_S(\theta)}$$

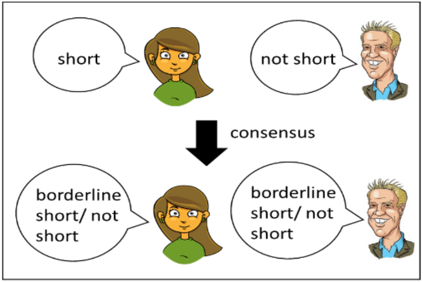
Question

Is it better to assert a sentence about which the sender is completely uncertain, or one which they are certain is borderline? Suppose that $\mu_S(\varphi) = 0.5$ *and* $\underline{\mu}_S(\theta) = 0$ *and* $\overline{\mu}_S(\theta) = 1$ *then* θ *is asserted if and only if*

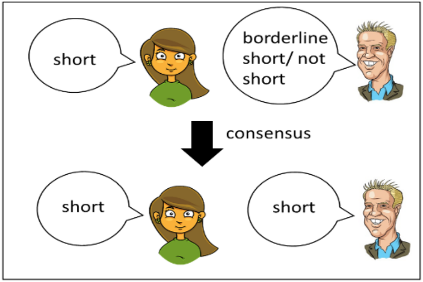
$$\alpha \leq 1 - \frac{1}{2}\frac{INF(\varphi)}{INF_2(\theta)}$$

Vagueness as a Route to Consensus

Can vagueness provide a middle ground when forming a consensus between individuals with different opinions?



Meeting in the middle



The strong opinion dominates

Combining and Comparing Three-valued opinions

- Suppose that we have a population of agents with opinions characterised by three-valued valuations.
- Agents with different opinions $\mathbf{v}_1$ and $\mathbf{v}_2$ both adopt a compromise position $\mathbf{v}_1 \odot \mathbf{v}_2$ according to the following truth table:

$\odot$	1	$\frac{1}{2}$	0
1	1	1	$\frac{1}{2}$
$\frac{1}{2}$	1	$\frac{1}{2}$	0
0	$\frac{1}{2}$	0	0

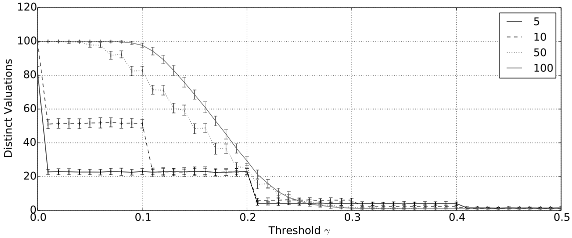
- Inconsistency: $I(\mathbf{v_1}, \mathbf{v_2}) =$

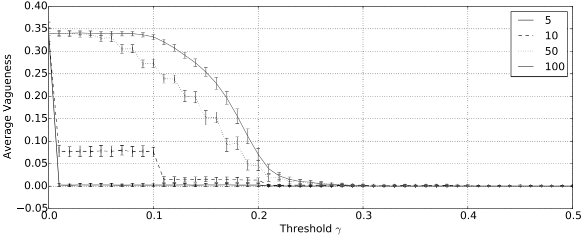
$$\frac{1}{n}|\{p_i : \mathbf{v}_1(p_i) = 1, \mathbf{v}_1(p_i) = 0 \text{ or } \mathbf{v}_1(p_i) = 0, \mathbf{v}_1(p_i) = 1\}|$$

- Vagueness: $V(\mathbf{v}) = \frac{1}{n}|\{p_i : \mathbf{v}(p_i) = \frac{1}{2}\}|$

Convergence to Crisp(er) Opinions

- Simulation: 100 agents, $50,000$ iterations
- Language: 5, 10, 50 and 100 propositions.
- Random initialisation of opinions.
- Bounded Confidence: operator applied if $I(\mathbf{v}_1, \mathbf{v}_2) \leq \gamma \in [0, 1]$.





Kilobot Decision Making

Kilobots are small coin-sized autonomous robots that can sense colour patterns in their environment, communicate locally to robots within range, control their motion through vibrating motors, and feedback their state using an LED.

Uncertain and Vague Opinions

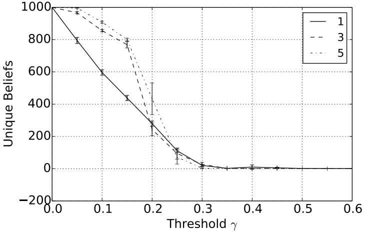
- Suppose that agents' opinions take the form of lower and upper measures $(\underline{\mu}(p_i), \overline{\mu}(p_i))$.
- We then extend the combination operator assuming that different agents hold independent opinions.

$\odot$	$1 : \underline{\mu}_1$	$\frac{1}{2} : \overline{\mu}_1 - \underline{\mu}_1$	$1 - \overline{\mu}_1$
$1 : \underline{\mu}_2$	$1 : \underline{\mu}_1\underline{\mu}_2$	$1 : (\overline{\mu}_1 - \underline{\mu}_1)\underline{\mu}_2$	$\frac{1}{2} : (1 - \overline{\mu}_1)\underline{\mu}_2$
$\frac{1}{2} : \overline{\mu}_2 - \underline{\mu}_2$	$1 : \underline{\mu}_1(\overline{\mu}_2 - \underline{\mu}_2)$	$\frac{1}{2} : (\overline{\mu}_1 - \underline{\mu}_1)(\overline{\mu}_2 - \underline{\mu}_2)$	$0 : (1 - \overline{\mu}_1)(\overline{\mu}_2 - \underline{\mu}_2)$
$0 : 1 - \overline{\mu}_2$	$\frac{1}{2} : \underline{\mu}_1(1 - \overline{\mu}_2)$	$0 : (\overline{\mu}_1 - \underline{\mu}_1)(1 - \overline{\mu}_2)$	$0 : (1 - \underline{\mu}_1)(1 - \overline{\mu}_2)$

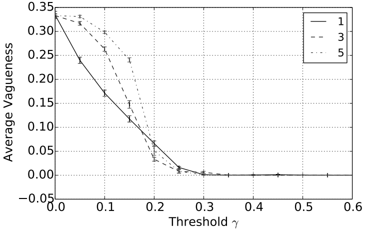
- This gives us the following lower and upper measures:

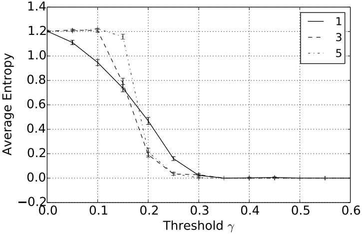
$$\underline{\mu}_1 \odot \underline{\mu}_2 = \underline{\mu}_1 \times \overline{\mu}_2 + \overline{\mu}_1 \times \underline{\mu}_2 - \underline{\mu}_1 \times \underline{\mu}_2$$
$$\overline{\mu}_1 \odot \overline{\mu}_2 = \underline{\mu}_1 + \underline{\mu}_2 + \overline{\mu}_1 \times \overline{\mu}_2 - \overline{\mu}_1 \times \underline{\mu}_2 - \underline{\mu}_1 \times \overline{\mu}_2$$

Convergence to Crisp Certainty



- We can think of this rule as a combination function $c : [0,1]^2 \times [0,1]^2 \to [0,1]^2$.
- Stable fixed points at $(0,0)$ and $(1,1)$.
- Unstable fixed point at $(\frac{1}{3}, \frac{2}{3})$.





Conclusion

You have to have an idea of what you are going to do, but it should be a vague idea. (Pablo Picasso)

- We have explored some possible ways in which vagueness might play a positive role in communication.
- This includes, vagueness as a source of stochasticity in channels with non-collaborating multiple senders.
- As a means of mitigating the risk of making forecast.
- Or as a route to consensus building between individuals with conflicting opinions.
- Different aspects of vagueness are useful in these cases.
- We have begun to explore how a more focussed investigation can inspire applications to artificial intelligence.
- Beyond vagueness there is perhaps a more general question about the role of more flexible conceptual models.