The limited expressiveness of single probability measures

Didier Dubois
Emeritus CNRS researcher
IRIT, Université Paul Sabatier
Toulouse, France

UNCERTAINTY :
representing graded belief.

- AN AGENT IS UNCERTAIN ABOUT A PROPOSITION IF (S)HE DOES NOT KNOW ITS TRUTH VALUE
  - **Examples**
    - The **probability** that the trip is more than one hour long is 0.7.
    - It is quite **possible** it snows to-morrow.
    - The agent has no **certainty** that Jean comes to the meeting
- HOW TO EVALUATE THE PROBABILITY, THE POSSIBILITY, THE CERTAINTY, THAT A PROPOSITION IS TRUE OR FALSE

Origins of uncertainty

- The variability of natural phenomena : **randomness.**
  - Coins, dice…: what about the outcome of the next throw?
- The lack of information: **incompleteness**
  - because of information is often lacking, knowledge about issues of interest is generally not perfect.
- Conflicting testimonies or reports: **inconsistency**
  - The more sources, the more likely the inconsistency

Probability Representations (on finite sets)

- A finite set S with n elements: A probability measure is characterized by a set of non negative weights $p_1, \ldots, p_n$, such that $\sum_{i=1,n} p_i = 1$.
  - p_i = probability of state s_i
- **Possible meanings of a degree of probability**:
  - Counting *favourable cases* for s_i over the number of possible cases assuming symmetry (coins, dice, cards)
  - *Frequencies from statistical information*: p_i = limit frequency of occurrence of s_i (**Objective probabilities**)
  - *Money involved in a betting scheme* (**Subjective probabilities)**

Remarks on using a single probability distribution

- **Computationally simple** : $P(A) = \sum_{s \in A} p(s)$
- **Conventions**: P(A) = 0 iff A impossible;

  P(A) = 1 iff A is certain;

  Usually P(A) = 1/2 for ignorance
- **Meaning :**
  - Objective probability is generic knowledge (statistics from a population)
  - Subjective probability on singular events (degrees of belief)

The two roles of probability

Probability theory is generally used for representing uncertainty due to the two types of issues:

1. **Randomness:** capturing variability through repeated observations.
2. **Partial knowledge:** representing belief in the face of information defect.

Note : these two issues are not mutually exclusive.

Measuring beliefs

Probability theory for uncertainty whatever its origin

1. **Frequencies** capture variability (**Hacking principle)**

Degrees of belief on n+1th trial outcome are equated to frequencies of the n previous observations of a repeatable phenomenon: P(A) = F(A)

2. **Belief in unique events** due to lack of information
   - via betting on lottery tickets for non-repeatable events
   - by analogical reasoning using thought frequentist experiment (balls in an urn)

SUBJECTIVE PROBABILITIES (Bruno de Finetti, 1935)

- p_i = *belief degree* of an agent on the occurrence of s_i
- measured as the price of a lottery ticket with reward 1 € if state is s_i in a betting game
- **Rules of the game**:
  - Banker sells tickets; gambler proposes prices p_i
  - They exchange roles if price p_i is too low
- **Why a belief state is a single distribution** ($\sum_j p_j = 1$):
  - Assume player buys all lottery tickets i = 1, …m.
  - If state s_j is observed, the gambler gain is $1 - \sum_j p_j$ and $\sum_j p_j - 1$ for the banker
  - *if* $\sum p_j > 1$ gambler *always loses money ;*
  - *if* $\sum p_j < 1$ *banker exchanges roles with* gambler
  - Only $\sum_j p_j = 1$ is rational

Bayesian probability

- **Bayesian postulate** : any state of knowledge should be represented by a single probability distribution:
  - Either via an exchangeable betting procedure
  - Or by using frequencies (real or thought ones)
- Not to do it is considered to be irrational (sure money loss, Dutch book argument)

What is the expressive power of probability distributions

Consequence of the Bayesian credo: in case of ignorance one is bound to use a uniform distribution.

But

Do uniform distributions represent ignorance ?

1. **Ambiguity :** do uniform bets express knowledge of randomness or plain ignorance?
2. **Instability :** the shape of a probability distribution is not scale-invariant, while ignorance is.
3. **Empirical falsification**: When information is missing, decision-makers do not always choose according to a single subjective probability (Ellsberg paradox).

Laplace principle of insufficient reason

- What is EQUIPOSSIBLE must be EQUIPROBABLE

- He states the problem in such a way make the elementary events equiprobable

  - *Argument of preserved symmetry*
  - *Also justified by the principle of maximal entropy*

Hence it is easy to believe that uniform distributions represent ignorance

Single distributions do not distinguish between incompleteness and variability

- VARIABILITY: Precisely observed random observations
- INCOMPLETENESS: Missing information
- **Example:** uniform probability on facets of a die
  - *A fair die tested many times:* Values are known to be equiprobable
  - *A new die never tested:* No argument in favour of a hypothesis against other ones, but frequencies are unknown.
- *BOTH CASES LEAD TO TOTAL INDETERMINACY ABOUT THE NEXT THROW (→ uniform distribution)*
- *BUT THEY DIFFER AS TO THE QUANTITY OF AVAILABLE INFORMATION*

The instability of uniform probabilistic representations of ignorance

- Suppose different domains U1 and U2 are used to describe the same problem (e.g. different vocabularies)
- So there is a most refined state space U and different one-to-many maps from U to U1 and U2
- **Claim**: a uniform probability distribution on U1 is generally not compatible with a uniform probability on U2.
  - This is natural if the distributions represent frequencies.
  - This is paradoxical if ignorance is represented by a uniform distribution

THE PARADOX OF IGNORANCE: finite case

- Case 1: life outside earth/ no life
  - ignorant's response 1/2 1/2
- Case 2: Animal life / vegetal only/ no life
  - ignorant's response 1/3 1/3 1/3
- They are inconsistent answers:
  - case 1 from case 2 : P(life) = 2/3 > P(no life)
  - case 2 from case 1: P(Animal life) = 1/4 < P(no life)
- **ignorance produces information *!!!!!***
- ***Uniform probabilities on distinct representations of the state space are inconsistent.***
- **Conclusion :** *a probability distribution cannot model incompleteness*

THE PARADOX OF IGNORANCE: infinite case

You have the same knowledge about x >0 as about y = f(x) (f bijection non linear such as 1/x, or Logx…).

- *x in [a, b] is equivalent to 1/x in [1/b, 1/a]*
- *But a uniform distribution on [a, b] is incompatible with a uniform distribution on [1/b, 1/a] : no scale invariance!*

Conclusion: uniform probability distributions do not represent ignorance.

(*It does not apply to frequentist distributions)*

LIMITATIONS OF BAYESIAN PROBABILITY FOR THE REPRESENTATION OF IGNORANCE

- ***Ignorance****: identical belief in any event different from the sure or the impossible ones*

- *A single probability cannot represent ignorance*: except on a 2-element set, the function g(A) = 1/2 $\forall A \neq S, \emptyset$, is NOT a probability measure.
- In the *life on other planets* example: 6 possible events that cannot have the same probability.

Ellsberg Paradox

- Savage claims that rational decision-makers choose according to expected utility with respect to a subjective probability
- **Counterexample**:An Urn containing
  - 1/3 red balls ($p_R = 1/3$)
  - 2/3 black or white balls ($p_W + p_B = 2/3$)
- For the ignorant Bayesian: $p_R = p_W = p_B = 1/3$.

- The game is to choose between games where you pick a ball and win or lose some money depending on the outcome.

- **Gambles should be prefered according to their Expected utility** :

$$u_a(R)p_R + u_a(W)p_W + u_a(B)p_B$$

based on a subjective probability dstribution.

Ellsberg Paradox

1. Choose between two bets

   B1: Win 1$ if red (1/3) and 0$ otherwise (2/3)

   B2: Win 1$ if white (≤ 2/3) and 0$ otherwise

   Most people prefer B1 to B2

2. Choose between two other bets (just add 1$ on Black)

   B3: Win 1$ if red or black (≥ 1/3) and 0$ if white

   B4: Win 1 $ if black or white (2/3) and 0$ if red (1/3)

   Most people prefer B4 to B3

But this is overwhelming empirical evidence that people make decisions in contradiction with utility theory based on a subjective probability

Ellsberg Paradox

- Let $0 < u(0) < u(1)$ be the utilities of gain.
- If decision is made according to a subjective probability assessment for red black and white: $(1/3, p_B, p_W)$:
  - B1 > B2:

    $EU(B1) = u(1)/3 + 2u(0)/3 > EU(B2) = u(0)/3 + u(1)p_w + u(0)p_B$
  - B4 > B3:

    $EU(B4) = u(0)/3 + 2u(1)/3 > EU(G) = u(1)\ (1/3 + p_B) + u(0)p_W$

  $\Rightarrow$ (summing, as $p_B + p_W = 2/3$) $2(u(0) + u(1))/3 > 2(u(0) + u(1))/3$: CONTRADICTION!
- Such an agent cannot reason with a unique probability distribution: **Violation of the sure thing principle.**

The sure thing principle

- An act ***a*** is a function from states *S* to consequences *X:*
  - If the state is $s \in S$ then consequence of ***a*** is $a(s) \in X$
  - **a1** ≥ **a2** iff EU(**a1**) ≥ EU(**a2**)
- *Ordering acts using expected utility satisfies the property that the preference of **a1** over **a2** does not depend on states where both acts have the same consequences.*
- Example:
  - **a1**(s) = 1 if s in A, 0 otherwise, then EU(**a1**) = P(A)
  - **a2**(s) = 1 if s in B, 0 otherwise then EU(**a2**) = P(B)
  - C disjoint from A and B
- STP: $A \geq B$ if and only if $A \cup C \geq B \cup C$

When information is missing, decision-makers do not always choose according to a single subjective probability

- *Plausible Explanation of Ellsberg paradox*: In the face of ignorance, the decision maker is pessimistic.
  - In the first choice, agent supposes $p_W = 0$: no white ball
    $EU(B1) = u(1)/3 + 2u(0)/3 > EU(B2) = u(0)$
  - In the 2d choice, agent supposes $p_B = 0$: no black ball
    $EU(B4) = u(0)/3 + 2u(1)/3 > EU(B3) = 2u(0)/3 + u(1)/3$

- **The agent does not use the same probability in both cases (**because of **pessimism):**
  - **the subjective probability depends on the proposed game.`**
  - **The epistemic state is a family of probability distributions**
  - **Ranking decisions by the lower expectation**

Summary on expressiveness limitations of subjective probability distributions

- The Bayesian dogma that any state of knowledge can be represented by a single probability is due to the exchangeable betting framework
  - Cannot distinguish randomness from a lack of knowledge in the computations.

- Representations by single probability distributions are language- (or scale-) sensitive

- When information is missing, decision-makers do not always choose according to a single subjective probability.

Main issue with single probability measures

- With a probability measure it is impossible to distinguish between
  - **Disbelief in** A (there is strong evidence against A)
  - **Lack of belief in** A (no evidence in favor of A)

  because $P(A^c) = 1 - P(A)$
- *Ignorance= no evidence for nor against A.*
- We need two set functions, one for certainty one for plausibility.

A GENERAL SETTING FOR REPRESENTING GRADED PLAUSIBILITY AND CERTAINTY

- 2 monotonic set-functions Pl and Cr from $\mathcal{E}$ to [0,1] called ***plausibility*** and ***certainty*** functions
  - generalize probability functions (Pl = Cr → P).
- **Conventions:**
  - Pl(A) = 0 "impossible" ;
  - Cr(A) = 1 "certain"
  - Pl(A) =1 ; Cr(A) = 0 "ignorance, Lack of belief "
    (no information)
  - Cr(A) ≤ Pl(A) "certain implies plausible"
  - Pl(A) = 1 − Cr(A^c) duality certain/plausible

How to represent partial ignorance?

- Using a subset of possible mutually exclusive values E for the variable x on S: « x in E »

  E is an epistemic state

- E can be a fuzzy set to express that some states are more possible than others
- Incomplete frequentist knowledge: epistemic state $\mathcal{P}$ on frequentist distributions P: *typically is a convex set of probabilities*

  (credal set)

How to represent belief ?

- Using a **credal set** $\mathcal{P}$: To each event A is attached a probability interval $[P_*(A), P^*(A)]$ such that
  - $Cr(A) = P_*(A) = \inf\{P(A), P \in \mathcal{P}\}$
  - $Pl(A) = P^*(A) = \sup\{P(A), P \in \mathcal{P}\} = 1 - P_*(A^c)$
- **Subjectivist interpretation** : $P_*(A)$ **is** a degree of belief measured by the maximal price for buying a lottery ticket
- with no exchangeability assumption (Walley).

$P^*(A)$ =minimal price for selling a lottery ticket

$\geq P_*(A)$

Special cases

- **Boolean necessity/possibility functions** based on epistemic state E

  $N(A) = 1$ if $E \subseteq A$, 0 otherwise (for belief)

  $\Pi(A) = 1 - N(A^c) = 1$ if $E \cap A \neq \emptyset$, 0 otherwise (for plausibility

  Represents a credal set $\mathcal{P} = \{P: P(E) = 1\}$

- **Graded necessity/possibility functions** based on fuzzy epistemic state E:

  $$\Pi(A) = \max_{s \in A} \pi(s); \qquad N(A) = 1 - \Pi(A^c)$$

  Represents a credal set $\mathcal{P} = \{P: P(A) \geq N(A)$ for all events $A\}$

- Using a **random epistemic state** (Dempster-Shafer), a probability distribution m over epistemic states:
  - $Bel(A) = \sum_{E_i \subseteq A,\, E_i \neq \emptyset} m(E_i)$ (expected necessity) $\qquad Pl(A) = 1 - Bel(A^c)$

  Represents a credal set $\mathcal{P} = \{P: P(A) \geq Bel(A)$ for all events $A\}$

Betting rates vs. States of Knowledge

- *Following Smets, one may distinguish two representation levels*
  - ***The credal level**: representing the belief state of the agent, accounting for partial ignorance (using belief functions)*
  - ***The betting level**: representing excnageable betting rates to form a probability function and compute expected utility.*

- *Betting rates are **induced** by belief states, but are **not in one-to-one correspondence** with them* : several states of knowledge may lead to the same betting rates.
  - *For instance, ignorance and randomness lead to uniform betting rates.*
- **One may want to derive a betting probability from a belief function**

Why not max entropy ?

- Suppose a person assesses belief that a coin falls on head (H) and tails (T).
- Cannot assess precise probabilities, only belief degrees as lower bounds
- Suppose he gives $Cr(H) = 0.4$, $Cr(T) = 0.1$ (lower probabilities), indicating a preference for H
- Maxent gives $P(H) = P(T) = 0.5$

When a credal set contains the uniform distribution, maxent always gives it.

It does not reflect the magnitudes of belief degrees.

Betting based on a belief function

- According to Smets
  - An agent has state of knowledge described by a mass function m.
  - The agent ranks decision using expected utility
- Generalized Laplace principle:
  - Select an epistemic state E with probability m(E)
  - Select an element at random in E (uniform on E)
- The betting probability used by the agent is

  $betp(s) = \sum\{m(E)/|E|, s \in E\}$
- It is the Shapley value of the belief function Bel, and the center of gravity of its credal set.

Maxent vs. Shapley value

- On the problem of Head vs. Tail assessments based on lower probabilities
  $Cr(H) = 0.4, Cr(T) = 0.1$:

  - Maxent : $Pr(H) = Pr(T) = 0.5$
  - Shapley value :

$Pr(H) = (Cr(H) + 1 - Cr(T))/2 = (0.4 + 1 - 0.1)/2 = 0.65$

$Pr(T) = 0.35$

Maxent vs. Shapley value

D. Dubois, A. Gilio and G. Kern-Isberner, Int. J. of Approx. Reasoning, 47(3): 333-351 (2008)

- Hypothesis H, piece of evidence E
- Suppose we know probabilities $P(E|H) = a$ and $P(E|H^c) = b$
- We do not have any prior probability on H.
- Credal set $\mathcal{P} = \{P: P(E|H) = a \text{ and } P(E|H^c) = b\}$
- How to compute the posterior $P(H|E)$?
  - Shapley value: $P(H|E) = a/(a+b)$ (like uniform prior)
  - Maxent: $P(H|E) = f(a)/(f(a)+f(b))$ with

    $f(x) = [x/(1-x)]^{(1-x)}$ Why ?????

SUBJECTIVE POSSIBILITY DISTRIBUTIONS

– *There are clearly several belief functions with a prescribed Shapley value P.*

- Consider the **least informative of those**, in the sense of a non-specificity index (expected cardinality of the random set) : $I(m) = \sum_{A \subseteq \Omega} m(A) \cdot card(A)$.
- Also the belief function having the least specific contour function $\pi_m(x) = \sum_{x \in E} m(E)$ among the isopignistic ones
- RESULT : The least informative belief function whose Shapley value is P *is unique and consonant.*

SUBJECTIVE POSSIBILITY DISTRIBUTIONS

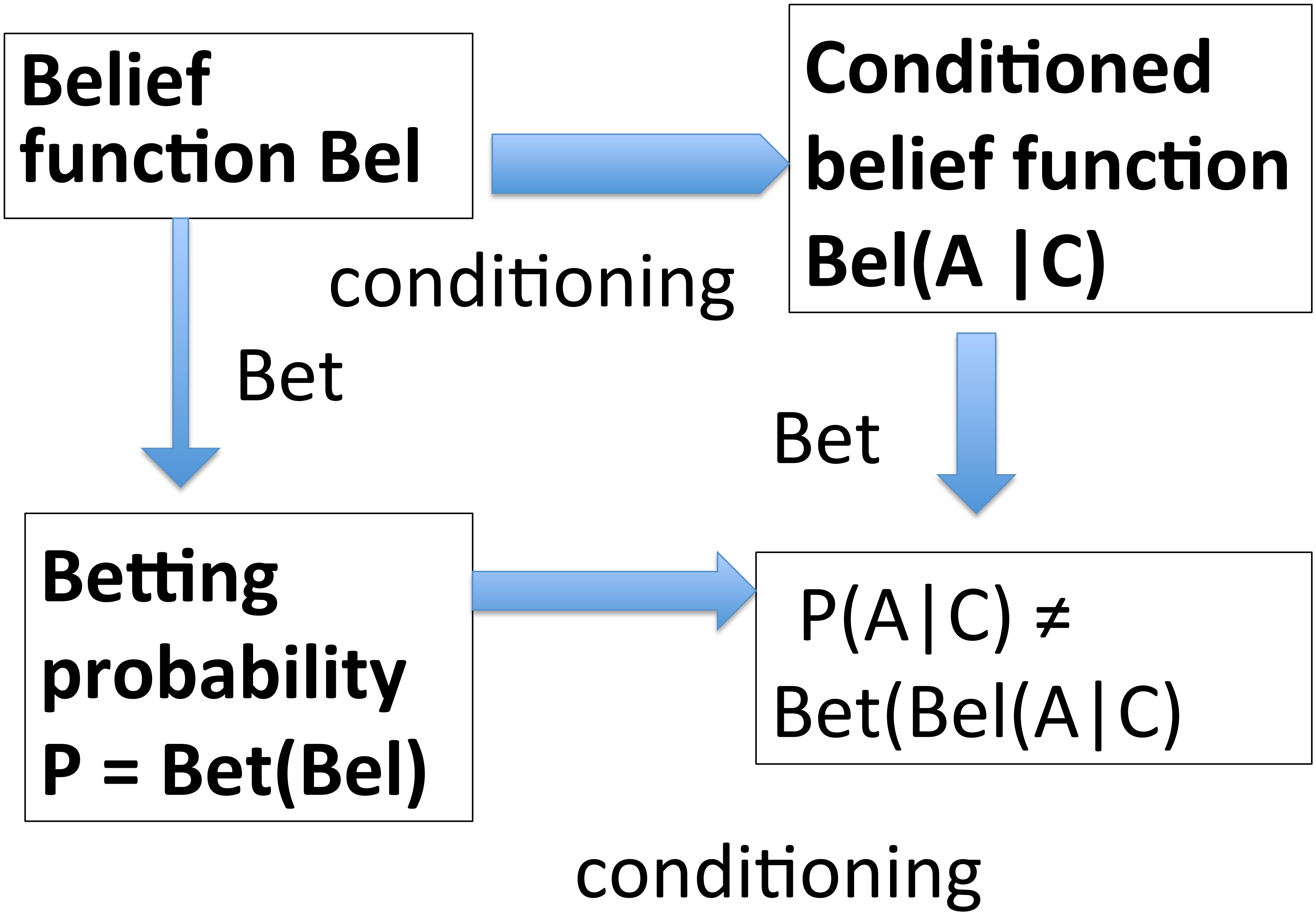
- The least specific belief function π^* in the sense of maximizing I(m) is characterized by

$$\pi^*_i = \Sigma_{j=i,\dots n} \min(p_j, p_i).$$

- It is a probability-possibility transformation, previously suggested in 1983: *This is the unique possibility distribution whose pignistic (Laplacean) probability is p.*

Revision: Credal vs. Betting levels

- Suppose a new sure information C is obtained
- Since betting rates cannot be equated with belief states, what should we revise?
  - conditioning at the credal level, and next, produce new betting rates ?
  - conditioning the previous betting rates ?


Belief function Bel
conditioning
Conditioned belief function Bel(A |C)
Bet
Bet
Betting probability P = Bet(Bel)
P(A|C) ≠ Bet(Bel(A|C)
conditioning

EXAMPLE OF REVISION OF EVIDENCE : The criminal case

- **Evidence 1** : three suspects : Peter Paul Mary
- **Evidence 2** : The killer was randomly selected man vs.woman by coin tossing.
  - So, S = { Peter, Paul, Mary}
- TBM modeling: The mass function is

  m({Peter, Paul}) = 1/2 ; m({Mary}) = 1/2
  - Bel(Paul) = Bel(Peter) = 0. Pl(Paul) = Pl(Peter) = 1/2
  - Bel(Mary) = Pl(Mary) = 1/2
- **Bayesian Modeling:** A prior probability
  - P(Paul) = P(Peter) = 1/4; P(Mary) = 1/2

- **Evidence 3** : Peter was seen elsewhere at the time of the killing.
- **TBM**: So Pl(Peter) = 0.
  - m({~~Peter~~, Paul}) = 1/2; m_t({Mary}) = 1/2
  - *A uniform probability on {Paul, Mary} results.*
- **Bayesian Modeling:**
  - P(Paul | not Peter) = 1/3; P(Mary | not Peter) = 2/3.
  - A very debatable result that depends on where the story starts. *Starting with i males and j females:*
    - P(Paul | Paul OR Mary) = j/(i + j);
    - P(Mary | Paul OR Mary) = i/(i + j)
- *Walley conditioning:*
  - Bel(Paul) = 0; Pl(Paul) = 1/2
  - Bel(Mary) = 1/2; Pl(Mary) = 1

Conclusion

- Single probability distributions do not properly reflect partial ignorance
  - Uncertainty theories extend probability theory for a more faithful/expressive representation of uncertainty
- Modelling and measuring the impact of ignorance is useful to trigger information collection decisions.
- Uncertainty theories allow for classical decision criteria via betting rates induced by epistemic states
  - Shapley value better than maxent.
- Other decision criteria can be used (lower expectation, generalizations of Hurwicz, etc.)