A Bayesian Interpretation of the Monty Hall Problem with Epistemic Uncertainty

Cristina Manfredotti, Paolo Viappiani

A Bayesian Interpretation of the Monty Hall Problem with Epistemic Uncertainty

Cristina Manfredotti[1] and Paolo Viappiani[2]

[1] UMR MIA-Paris, AgroParisTech, INRA, University of Paris-Saclay, 75005 Paris, France
cristina.manfredotti@agroparistech.fr
[2] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
paolo.viappiani@lip6.fr

Abstract. The Monty hall problem is a classic puzzle that, in addition to intriguing the general public, has stimulated research into the foundations of reasoning about uncertainty. A key insight to understanding the Monty Hall problem is to realize that the specification of the behavior of the host (i.e. Monty) of the game is fundamental. Here we go one step further and reason, in Bayesian way, in terms of epistemic uncertainty about the behavior of host, assuming subjective probabilities.
We also consider several generalizations of the classic Monty hall problem considering different priors for the doors, several doors instead of three, and different ways the host can choose which door to open when several are possible. We show that in these generalized versions, the player faces a sequential decision problem, since the choice of the first door is key. We provide a general solution for the most general case using decision trees and determine the optimal policy.

1 Introduction

The Monty hall problem [12–14] is a classic puzzle that, in addition to intriguing the general public, has stimulated research [1, 2] into the foundations of reasoning about uncertainty. It is stated as follows:

> Suppose you're on a game show, and you're given the choice of three doors: Behind one door is a car; behind the others, goats. You pick a door, say No. 1, and the host, who knows what's behind the doors, opens another door, say No. 3, which has a goat. He then says to you, "Do you want to pick door No. 2?" Is it to your advantage to switch your choice?

The commonly accepted answer is that it is best to switch. Indeed, assuming that the prize is placed behind a door according to a uniform distribution, by choosing to switch the player obtains probability $\frac{2}{3}$ of getting the prize.

This is true however under a particular assumption about the behavior of the host: the host always opens a door; this door is different than the one that the player has chosen and from the one with the prize behind it. Indeed, several authors have argued [10, 1, 2] that the answer to the puzzle crucially depends on the behavior of the host.

In this paper we go one step further and consider uncertainty over which protocol Monty might be following. We reason about Monty's behavior using subjective probabilities about the possible protocols; therefore *we move from representing uncertainty over the placement of the doors, to representing our epistemic uncertainty over the behavior of the host*. Moreover, we consider some generalizations of the Monty Hall problem supposing that the position of the car might be not distributed uniformly. When considering these generalized settings, we realize that the solution to the problem is a policy dictating which door should we choose at each step of the game. While the Monty Hall problem has been extensively studied before in the computer science and applied mathematics literature [1, 5, 7, 6, 10, 11, 15], we do not know any works that consider the generalized settings that we address here.

2 Epistemic Uncertainty over Monty's Protocol

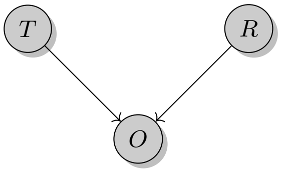

Fig. 1. A Bayesian network formalizing the Monty Hall problem with uncertainty over the host's protocol.

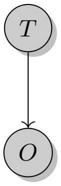

Fig. 2. Simplified Bayesian network for the Monty Hall problem. The uncertainty over Monty's protocol is now integrated in the conditional probability $P(O|T)$.

In this Section we consider the Monty Hall Problem (MHP) with 3 doors and we explicitly reason in terms of epistemic uncertainty about the host's (i.e. Monty's) behavior. We assume for the moment that the car is equally likely to be behind any of the doors. Different assumptions about the host's protocol can be made:

- *AO (always open)*: this is the "classic" Monty's behavior. The host always opens a door that has a goat behind it and hasn't been picked by the player (if the player initially picked the door with the car, then he randomly chooses one of the two other doors);
- *RO (open at random)*: Monty randomly chooses a door and, if there is no car behind it and it has not been picked by the player, then open it, while no door is opened if the randomly chosen door hides the car;

- *SO (selective open)*: the choice of opening a door depends on specific conditions (whether the player picked the door with the prize). In particular we consider "benevolent" Monty (opens a door whenever the player is pointing at a door with a goat, and not when the player picked the door with the car; this behavior is dubbed SO_+) and "adversarial" Monty (opens a door only when the player is pointing at the car; SO_-)

While under the AO protocol the player has an advantage to switch, under RO, switching gives no advantage, as it has been noticed several times (see, for instance, exercise 3.9 in the book of MacKay [8] on page 57 and its solution on page 61). Obviously under SO_+ it is always beneficial to switch and under SO_- one should never switch (see also Halpern's book [2] on pages 216-217).

We now assume now that Monty's behavior is a situation of epistemic uncertainty: the player does not know exactly which protocol Monty has adopted and this uncertainty is represented by a probability distribution. This means that, from the point of view of the player, Monty is behaving according to a mixture of the protocols above. This mixture is given by the parameters $\boldsymbol{\theta} = (\theta_{AO}$, θ_{RO}, θ_{SO_+}, $\theta_{SO_-})$, where θ_{AO} is the probability of adopting the AO protocol, and so on; in other words $\boldsymbol{\theta}$ is the subjective probability distribution of Monty's behavior. Actually, our model allows for the possibility that Monty itself is behaving according to a mixture of the protocols, but the player has no access to the true mixture parameters and makes use of subjective probabilities instead.[3]

We formalize the Monty hall problem using the Bayesian network depicted in Figure 1 with three nodes: T, R, and O. Node T represents the event "the player has pointed to the door with the car behind", R takes value in $\mathcal{R} = \{AO, RO, SO_+, SO_-\}$, that is the set of possible protocols. O is the event "Monty opened a door".

Assuming that the car is uniformly distributed between the three positions, we write the values of the Conditional Probability Tables (CPTs) for the nodes of the Bayesian network. For node T we have:

$$P(T) = \frac{1}{3} \qquad\qquad P(\neg T) = \frac{2}{3}$$

and for R:

$$P(R) = \theta_R \quad \forall R \in \{AO, RO, SO_+, SO_-\}.$$

We now write the probability of the event O (the host opens a door) conditioned on T (the player has pointed at door with the car) and on the protocol. These are the CPT values associated with the node O in Figure 1.

$$\begin{aligned} &P(O|T, AO) = 1 &\qquad& P(O|\neg T, AO) = 1 \\ &P(O|T, RO) = 0.66 && P(O|\neg T, RO) = 0.33 \\ &P(O|T, SO_+) = 0 && P(O|\neg T, SO_+) = 1 \\ &P(O|T, SO_-) = 1 && P(O|\neg T, SO_-) = 0 \end{aligned}$$

[3] A possible extension of this work could investigate the use of Bayesian hierarchical models, adopting prior distributions on the mixture's parameters.

From the belief $\boldsymbol{\theta}$ we can determine the probability of the host opening a door (event O) given the initially chosen door conceals the car (T) or the goat ($\neg T$):

$$P(O|T) = \sum_{r \in \mathcal{R}} \theta_r P(O|T, r) = \theta_{AO} + \frac{2}{3}\theta_{RO} + \theta_{SO_-} \tag{1}$$

$$P(O|\neg T) = \sum_{r \in \mathcal{R}} \theta_r P(O|\neg T, r) = \theta_{AO} + \frac{1}{3}\theta_{RO} + \theta_{SO_+} \tag{2}$$

The above equations allow us to reduce our problem to the simplified Bayesian network given in Figure 2 (where $\boldsymbol{\theta}$ can be seen as a vector of parameters).

Using basic probability calculus and Bayes theorem, we can derive a condition on $\boldsymbol{\theta}$ for when switching is advantageous. We compute the probability (from the point of view of the player) that the car is behind the initially picked door conditioned to observing that the host has opened another door, using Bayes theorem:

$$P(T|O) = \frac{P(O|T)P(T)}{P(O)} = \frac{P(O|T)P(T)}{P(O|T)P(T) + P(O|\neg T)P(\neg T)} = \frac{P(O|T)}{P(O|T) + 2P(O|\neg T)}$$

If the player sticks to his initial guess, then $P(T|O)$ is the probability of getting the car. If the player switches, the car is found with probability $P(\neg T|O) = 1 - P(T|O)$. Switching is then advantageous when

$$P(\neg T|O) > P(T|O) \iff P(O|\neg T)P(\neg T) > P(O|T)P(T) \tag{3}$$

$$\iff P(O|\neg T) > \frac{1}{2}P(O|T). \tag{4}$$

Since we want to know under what condition with respect to $\boldsymbol{\theta}$ switching is advantageous, we now expand the expression above using Equations (1) and (2):

$$\frac{2}{3}(\theta_{AO} + \frac{1}{3}\theta_{RO} + \theta_{SO^+}) > \frac{1}{3}(\theta_{AO} + \frac{2}{3}\theta_{RO} + \theta_{SO^-}) \tag{5}$$

$$\iff \frac{1}{3}\theta_{AO} + \frac{2}{3}\theta_{SO^+} - \frac{1}{3}\theta_{SO^-} > 0 \tag{6}$$

We note that in the computation just above, we were only interested in determining when switching is beneficial[4]; that is, we did not considered the situations in which no door is opened, and no choice if offered. Considering a game episode starting with the initial door selection, we are now interested in computing the total expected payoff of the two policies “switch” (switch door when possible) and “keep”, where we define payoff as 1 if the player gets the car at the end of the game, and 0 otherwise. Note that the two policies imply the same outcome when Monty does not open a door (and therefore does not offer the possibility to switch choice).

- The policy “keep” obviously achieves expected payoff $\frac{1}{3}$.

[4] Indeed the original statement of the MHP concerns the specific decision of what to do when offered the possibility of switching.

– The policy “switch” achieves expected payoff

$$P(T, \neg O) + P(\neg T, O) = P(T)P(\neg O|T) + P(\neg T)P(O|\neg T)$$

since the car is won if the player initially picked the right door and the host *does not* open any door (there is no option to switch, and therefore the car is obtained), or if the initial guess is wrong but the host *does* open a door thus offering the chance to switch (the offer is then accepted, since we're following the “switch” policy, and the car is obtained). Therefore, since $P(T) = \frac{1}{3}$, the payoff of the “switch” policy is $\frac{1}{3} - \frac{1}{3}P(O|T) + \frac{2}{3}P(O|\neg T)$ or, equivalently, $\frac{1}{3} + \frac{1}{3}\theta_{AO} - \frac{1}{3}\theta_{SO_-} + \frac{2}{3}\theta_{SO_+}$.

We observe that in the 3-doors setting, with Monty uniformly random when the player chooses the door with the car in the first round, there are really just two parameters: $P(O|T)$ and $P(O|\neg T)$. Given these two values, the distribution over the protocol R is identified according to Equations (1) and (2); note however that different $\boldsymbol{\theta}$ may project to the same $P(O|T)$ and $P(O|\neg T)$ values.

The following proposition summarizes our analysis:

Proposition 1. *The payoffs of the two policies “keep” and “switch” are:*

$$V(keep) = \frac{1}{3}$$
$$V(switch) = \frac{1}{3} - \frac{1}{3}P(O|T) + \frac{2}{3}P(O|\neg T) = \frac{1}{3} + \frac{1}{3}\theta_{AO} - \frac{1}{3}\theta_{SO_-} + \frac{2}{3}\theta_{SO_+}$$

Switching is advantageous when $P(O|T) < 2P(O|\neg T)$, or equivalently, when $\frac{1}{3}\theta_{AO} + \frac{2}{3}\theta_{SO^+} - \frac{1}{3}\theta_{SO^-} > 0$.

Example 1. Assume the player is not given any information about the host's behavior. The player reasons that the host might be following one of the four protocols AO, RO, SO$_+$ and SO$_-$. In absence of any prior information, a reasonable way for the player to proceed is to consider a uniform prior on the host's protocol: with $\boldsymbol{\theta} = (0.25, 0.25, 0.25, 0.25)$, thus we have that $\frac{1}{3}\theta_{AO} + \frac{2}{3}\theta_{SO^+} - \frac{1}{3}\theta_{SO^-} = \frac{1}{6} > 0$, so switching is advantageous according to Proposition 1.

Another reasonable uninformative prior is to suppose $P(O|T) = P(O|\neg T) = 0.5$; this also means that switching is advantageous.

Example 2. Assume now that the player has access to the history of past behaviors of the host in n previous games. The player can use Laplace's rule (equivalent to assuming a Beta prior) to estimate the probability of opening a door. Let n_T the number of episodes where the initially picked door hid the car; $n = n_T + n_{\neg T}$. Let o_T be the number of observations consisting in the host opening a door when the initially picked one is correct. The player estimates the probabilities: $\hat{p}_{O|T} = \frac{o_T+1}{n_T+2}$ and $\hat{p}_{O|\neg T} = \frac{o_{\neg T}+1}{n_{\neg T}+2}$. Equation 4 is then used with these estimations to decide whether to switch or not.

3 Different priors for doors

We now consider the situation where each door is associated with a prior probability p_i of concealing the prize (for short we will use the term probability of a door). We still consider that there are three doors and delay the extension to an arbitrary number of doors to Section 4.

Unlike the original statement of the puzzle, the choice of the first door is critical (since doors cannot anymore treated as indistinguishable). The behavior of the player is fully specified by a decision policy; with 3 doors a policy is a pair (i, a), where i is the index of a door and a is either "switch" or "keep" (in case Monty offers such possibility). We still assume that the host is not biased, in the sense that, if the player initially picks the door with car, the host, if he decides to open a door, is just as likely to open any of the two remaining doors.

We now compute the expected payoff V of the strategy (i, switch) using the observation that the car is obtained in two cases i) if the door i conceals the car and the host does not open a door (and so he does not offer to switch) and ii) if the door i does not conceal the car and the host does offer to switch; hence:

$$\begin{aligned}
V(i, \text{switch}) &= P(\neg O, T) + P(O, \neg T) \\
&= P(\neg O|T)p_i + P(O|\neg T)(1 - p_i) \\
&= (1 - \alpha_T)p_i + \alpha_{\neg T}(1 - p_i) \\
&= (1 - \alpha_T - \alpha_{\neg T})p_i + \alpha_{\neg T}
\end{aligned}$$

where we let $\alpha_T := P(O|T)$ and $\alpha_{\neg T} := P(O|\neg T)$. On the other hand, the payoff of strategy (i, keep) is obviously p_i :

$$V(i, \text{keep}) = p_i.$$

The following inequality gives the condition that makes switching beneficial.

$$\begin{aligned}
V(i, \text{switch}) > V(i, \text{keep}) &\iff (1 - \alpha_T - \alpha_{\neg T})p_i + \alpha_{\neg T} > p_i && (7) \\
&\iff p_i < \frac{\alpha_{\neg T}}{\alpha_T + \alpha_{\neg T}} && (8)
\end{aligned}$$

Equation (8) provides a condition to check to determine whether (i, keep) or (i, switch) is best. However, in order to identify the best policy, we need to account as well the choice of i, i.e. the first door. There are 6 possible policies, but in fact some are dominated: among the "keep" policies, the best one is to pick the door i^+ associated with highest prior $p^+ = \max_{i \in \{1,2,3\}} p_i$. On the other hand, if we switch, it is not so obvious if the initial choice should be a door with high or with low prior probability. We therefore consider several different cases.

- If $\alpha_T + \alpha_{\neg T} > 1$ then the payoff $V(i, \text{switch})$ decreases when p_i increases; hence among all policies that switch door in the second step, the best one is to pick, in the first round, the door with lowest prior probability.
  Therefore to determine the optimal policy we compare p^+ (the payoff of selecting the door with highest p and then keeping this choice) and (1 −

$\alpha_T - \alpha_{\neg T})p^- + \alpha_{\neg T}$, the value of the payoff obtained by picking door $i^- = \arg\min_i p_i$ and then switching. The condition to check is

$$(1 - \alpha_T - \alpha_{\neg T})p^- + \alpha_{\neg T} > p^+.$$

In other words: we pick i^- and switch in the case $p^+ + (\alpha_T + \alpha_{\neg T} - 1)p^- - \alpha_T < 0$; otherwise we pick i^+ and keep the same choice.

- If $\alpha_T + \alpha_{\neg T} = 1$ then, assuming that we switch, it does not matter which door we select initially: the payoff $V(i, \text{switch})$ will be always $\alpha_{\neg T}$ for any $i = 1, 2, 3$. Hence, if $p^+ > \alpha_{\neg T}$ we select the door with highest prior and keep this choice, and otherwise choose any door and switch.
- If $\alpha_T + \alpha_{\neg T} < 1$ then the payoff $V(i, \text{switch})$ increases when p_i increases. We therefore initially pick i^+, the door with highest prior, and we compare the payoff of either switching or keeping. Hence, if $p^+ > \frac{\alpha_T}{\alpha_T + \alpha_{\neg T}}$ then the optimal policy is (i^+, keep) otherwise it is (i^+, switch).

Proposition 2. *The payoff of the policies are as follows:*

$$V(i, \textit{keep}) = p_i$$
$$V(i, \textit{switch}) = (1 - \alpha_T - \alpha_{\neg T})p_i + \alpha_{\neg T}$$

Obviously this model generalizes that of the previous section. Indeed, if we substitute $p_i = \frac{1}{3}$ in Equation (8) we determine the condition $\frac{\alpha_T}{\alpha_{\neg T}} < 2$ for switching being advantageous, as shown in the previous section in Equation (4).

We now consider, as examples, two particular cases.

Example 3. Assume three doors with prior probability p_1, p_2, p_3 and that the host behaves according to the AO protocol of Section 2 (the player may know this having observed previous games), that means $\alpha_T = \alpha_{\neg T} = 1$. Now, if you pick door i initially, switching gives $1 - p_i$; keeping the same choice gives you p_i. The best policy is to pick the door with least value of the prior probability, wait for the host action and then switch door; the optimal payoff is:

$$V^* = 1 - \min_i p_i.$$

This value is strictly higher than the value of the policy of picking the door with highest p_i and not switching, unless $\max_i p_i = 1$.

Example 4. We now consider, as special case of the scenario studied in this section, that the host does not allow to switch with probability q, regardless of whether the player points at the right door or not; the host opens a door allowing to switch with probability $1 - q$. In other words $\alpha_T = \alpha_{\neg T} = 1 - q$. The payoff of (i, switch), the policy "pick door i and switch when offered", is then:

$$V(i, \text{switch}) = qp_i + (1 - q)(1 - p_i) = (2q - 1)p_i + 1 - q.$$

We analyze the different cases:

- If $q < 0.5$ and $p^+ \geq (2q - 1)p^- + 1 - q$ then (i^+, keep) is an optimal policy

- If $q < 0.5$ and $p^+ \leq (2q-1)p^- + 1 - q$ then (i^-, switch) is an optimal policy
- If $q = 0.5$ and $p^+ \leq 0.5$ then (i, switch), for all $i \in \{1,2,3\}$, are optimal policies
- If $q = 0.5$ and $p^+ \geq 0.5$ then (i^+, keep) is an optimal policy
- If $q > 0.5$ and $p^+ \geq 0.5$ then (i^+, keep) is an optimal policy
- If $q > 0.5$ and $p^+ \leq 0.5$ then (i^+, switch) is an optimal policy

For the last two cases, notice that when $q > 0.5$, the condition $V(i^+, \text{keep}) \geq V(i^+, \text{switch})$ simplifies to $p^+ \geq 0.5$.

4 General setting: *n* doors and general response model

In this section we analyze the general formulation of the MHP and develop a model based on sequential decision making. We consider the general situations with n doors and arbitrary prior probabilities p_i. Monty may decide not to open any door. Moreover, in this section we allow for Monty to be biased with respect to which door to open when he can choose among several unopened doors. Note that the models discussed in the previous Sections can be seen as special cases of this general model.

The problem is solved with a decision tree; an excerpt of the general tree is shown in Figure 3. Note that we use a different notation from previous Sections, since the generalized problem does not enjoy the symmetries that simplified the treatment of the former models. Each node of the tree is labeled with the variable (either a decision or a random variable) that it represents.

The decision node S represents the initial door choice, with possible choices in $\{S_1, \ldots, S_n\}$. For each S_j, there is a chance node O, with outcomes in $\{O_\emptyset\} \cup \{O_i\}_{i \neq j}$ representing whether and which door the host opens; in our notation $O_\emptyset$ is the event no door is opened, and O_i means that the door i is opened; then:

- In case no door is opened, $O = O_\emptyset$, the position of the car is revealed to be at a position k in a chance node T, with outcomes in $\{T_1, \ldots, T_n\}$. In the leaf nodes, utility is 1 if $j = k$, and 0 otherwise.
- If, instead, the event O_i happens, we face the decision node F, with possible choices in $\{F_l\}_{l \neq i}$, representing the final door choice[5], with the choice that must be different from i. Then, the chance node T, with outcomes in $\{T\}_{k \neq i}$, reveals the car's position; utility is 1 if the choice for node F is the same as the outcome of T.

Let $\alpha_{i,j,k} := P(O_i | S_j, T_k)$ to be the probability of Monty opening door i given player's selection of door j and car in door k. The vector

$$(\alpha_{i,j,k})_{i,j,k \in \{1,\ldots,n\}}$$

[5] In this generalized model, switching occurs when the choice at node F is a different door from the one chosen at the root S.

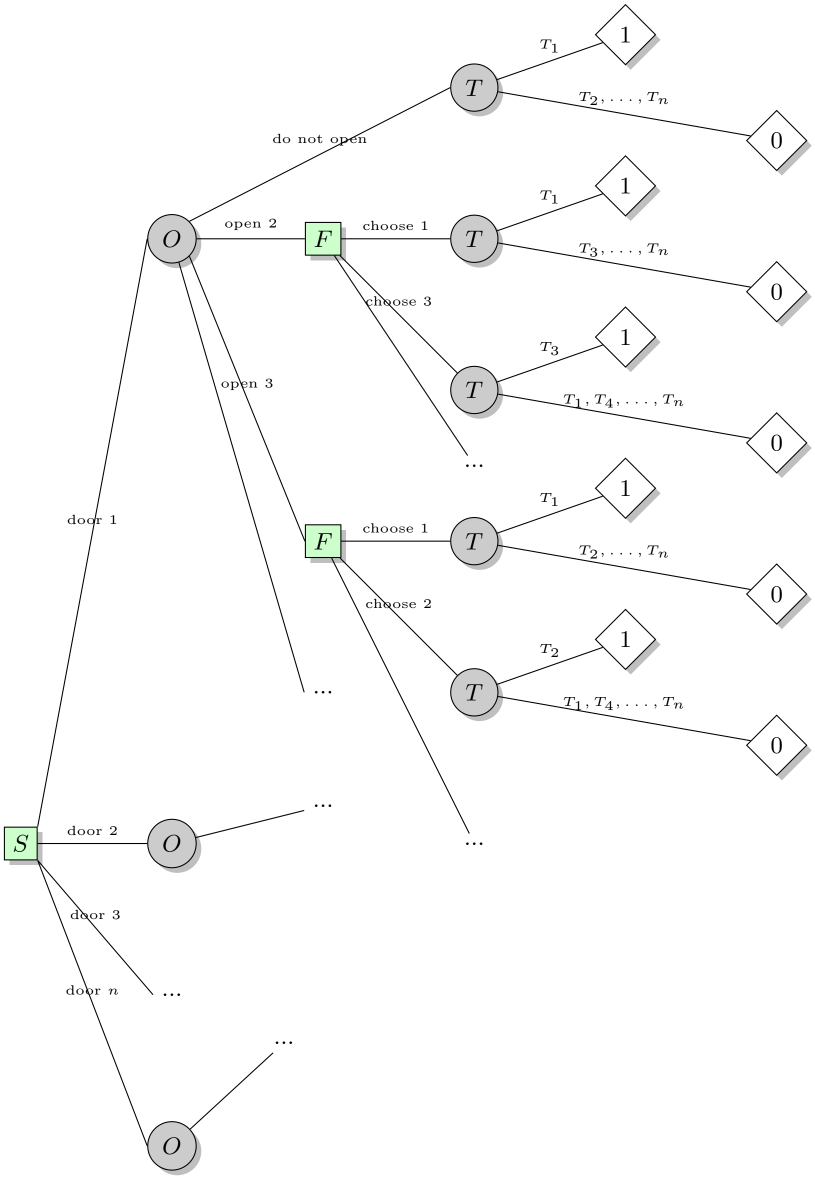

Fig. 3. The decision tree corresponding to the generalized Monty Hall problem. The root, the decision node S, is displayed on the left. In a chance node, the information available up to that point is used to condition the distribution; for example, if the player selected door 1 initially, the variable O is distributed according to $P(O|S_1)$, that can be computed using Equation 9. Similarly, if the player chooses door i and the host does not open any door, the probability of T_i is given by $P(T_i|S_i, O_\emptyset)$.

fully describes Monty's behavior from the point of view of the player. Note that because Monty never opens the door chosen by the player, we have $\alpha_{i,i,k} = 0$, and because Monty never opens the door with the car, we have $\alpha_{k,j,k} = 0$

Note that, in this decision tree, the probabilities associated to chance nodes represent the epistemic uncertainty of the player about the behavior of Monty and as well the position of the car.

A solution to a decision tree is a strategy that specifies how the player should act at the various decision nodes. The optimal strategy can be found by the "averaging out and fold back" method (see, for instance, the book of Jensen and Nielsen [4]). The probability of Monty opening one specific door i, when the player has picked door j, can be determined by marginalization:

$$P(O_i|S_j) = \sum_{k=1}^{n} P(O_i|S_j, T_k)P(T_k) = \sum_{k=1}^{n} \alpha_{i,j,k} p_k. \tag{9}$$

This means that, from the point of view of the player (that does not know where the car is located), Monty does not open any door with probability

$$\beta_j := 1 - \sum_{i=1}^{n} P(O_i|S_j) = 1 - \sum_{k=1}^{n}\sum_{i=1}^{n} \alpha_{i,j,k} p_k$$

when the player chooses door j initially.

We now solve the decision tree starting the evaluation from the nodes at the bottom. For the T nodes, we use Bayes in order to determine the posterior probability of each door. We determine the value $p'_k := P(T_k|O_i, S_j)$, the posterior probability of the car being placed behind door k after having observed that Monty opened door i and after having initially picked door j:

$$p'_k = P(T_k|O_i, S_j) = \frac{P(O_i|S_j, T_k)P(T_k|S_j)}{P(O_i|S_j)} = \frac{\alpha_{i,j,k} p_k}{\sum_{k'=1}^{n} \alpha_{i,j,k'} p_{k'}}$$

where we used $P(T_k|S_k) = P(T_k)$, since the selection of a door does not influence where the car lies.

At each of the F nodes, we need to choose the door with highest posterior p'_k given our initial choice S and the host's action. This means picking the door giving $\max_k p'_k = \max_k P(T_k|O_i, S_j)$.

At the O nodes, the host is acting. He might not open any door (probability β_j) or open a door i with probability $P(O_i|S_j)$.

- If the host is not opening any door, the player is successful only if the door with the car is the one that he initially picked. The probability of this is

$$P(T_j|O_\emptyset, S_j) = \frac{P(O_\emptyset|S_j, T_j)P(T_j)}{P(O_\emptyset|S_j)} = \frac{(1 - \sum_i \alpha_{i,j,j})p_j}{\beta_j}$$

  and the contribution to the O node is $P(T_j|O_\emptyset, S_j)$ times $P(O_\emptyset|S_j)$.
- If, instead, the host opens door i, the contribution to the value of the node is $P(O_i|S_j) \max_k P(T_k|O_i S_j)$.

This gives the following value for a node of type O:

$$P(O_\emptyset|S_j)P(T_j|O_\emptyset, S_j) + \sum_{i=1}^{n} P(O_i|S_j) \max_k P(T_k|O_i, S_j) = \tag{10}$$

$$\beta_j \frac{(1 - \sum_i \alpha_{i,j,j})p_j}{\beta_j} + \sum_{i=1}^{n} P(O_i|S_j) \max_k \frac{\alpha_{i,j,k}p_k}{P(O_i|S_j)} = \tag{11}$$

$$\Big(1 - \sum_{i=1}^{n} \alpha_{i,j,j}\Big)p_j + \sum_{i=1}^{n} \max_k \alpha_{i,j,k}p_k \tag{12}$$

At the root, we have the decision node S where we take the door j that maximizes the value of Equation (12).

Proposition 3. *The optimal policy achieves expected payoff:*

$$V^* = \max_{j=1,\dots,n} \Big[\Big(1 - \sum_{i=1}^{n} \alpha_{i,j,j}\Big)p_j + \sum_{i=1}^{n} \max_k \alpha_{i,j,k}p_k\Big].$$

Example 5. We now consider *classic Monty with response bias*, that is the scenario with 3 doors and uniform priors, $P(T_i) = \frac{1}{3}$, the host always open one door (AO protocol), but when the player chooses the door with the car behind in the first step, then the host may not be following an uniform distribution in deciding which door to open (see also [2], pages 216-217).

In the following description let i, j and k to be distinct; i.e. (i, j, k) is a permutation of $(1, 2, 3)$. We have $\alpha_{i,j,k} = 1$ and $\alpha_{i,j,j} + \alpha_{k,j,j} = 1$. Now, assume that the player selects door j and the host opens door i. Observe that the total probability of opening door i is $P(O_i|S_j) = \frac{1}{3}(1 + \alpha_{i,j,j})$. We then determine the posterior probabilities for positions j and k (the car cannot be behind door i since this door was opened): $P(T_k|S_j, O_i) = \frac{\alpha_{i,j,k}p_k}{\alpha_{i,j,j}p_j + \alpha_{i,j,k}p_k} = \frac{1}{\alpha_{i,j,j}+1}$ and $P(T_j|S_j, O_i) = \frac{\alpha_{i,j,j}p_j}{\alpha_{i,j,j}p_j + \alpha_{i,j,k}p_k} = \frac{\alpha_{i,j,j}}{\alpha_{i,j,j}+1}$. The best decision in the second stage of the game consist in picking the door j or k associated with the higher posterior. Now, consider the decision at the root. The payoff $V(S_j)$ of selecting door j, assuming that then choosing optimally in the second step, is:

$$V(S_j) = \sum_{i \neq j} \frac{\alpha_{i,j,j} + 1}{3} \max\{\frac{1}{\alpha_{i,j,j} + 1}, \frac{\alpha_{i,j,j}}{\alpha_{i,j,j} + 1}\} = \frac{1}{3} \sum_{i \neq j} \max\{1, \alpha_{i,j,k}\} = \frac{2}{3}$$

Since this value does not depend on j, the first door can be chosen in an arbitrary way. It turns out that the best policy in this case is "pick any door randomly and then, after that the host opens a door, switch choice to other unopened door". The optimal value of the optimal policy is $V^* = \frac{2}{3}$.

5 Discussion and Conclusions

The Monty Hall Problem (MHP) is a puzzle that has raised a lot of attention and is frequently used as a didactic tool for explaining how to reason with subjective probabilities. Some interesting variations of the Monty Hall problem have been analyzed by Lucas et al. [7, 6]; we refer the reader to the book of Rosenhouse that provide an excellent review of materials on the Monty Hall problem [11].

In computer science, the Monty hall problem has stimulated a variety of research activities, including works on epistemic logic [5] and reasoning about uncertainty [2]; we also mention the interpretation given by Viappiani and Boutilier (in the appendix of [16]) in terms of preferences and choice. On the other hand, psycologhists have used the Monty Hall problem to study how human people reason with probabilities [15].

In this paper we provided an analysis of the Monty hall problem and some of its extensions emphasizing the role of dealing with epistemic uncertainty. We have considered policies that determine which door to select in the first round, and whether to keep the same choice or to switch in the second. We provided the characterization of the optimal policy in several generalizations of the MHP: considering different prior subjective probabilities for the position of the prize behind the doors, considering uncertainty over the possible host's behaviors and considering n doors. We mention some interesting further extensions of the MHP worth studying: considering the generalization m rounds, and the case where the number of rounds is uncertain.

We now provide some brief comments on how the Monty Hall problem is related to several areas of artificial intelligence. First of all, notice that the tools we have used (Bayesian reasoning, Bayesian networks, and decision trees) are typically used in AI. Moreover, some of the ideas behind our work are relevant to research in *multi-agent systems* since agents often have to reason about other agents' behaviour. In some sense, the MHP can be seen as an emblematic case of an agent reasoning about another agent's behavior, a key aspect of multi agent system research; we advocate that it often worth to consider a wide variety of possible behaviors and not just a single one, and to consider mixture of such possible behaviors (as we did in our treatment of the MHP). This could be of relevance for opponent modeling in games, for instance.

The Monty hall problem has connections with the statistical areas of selectively reported data and missing data; in particular the missing at random hypothesis in machine learning [3]. In the case of *recommender systems* based on collaborative filtering where users rate items such as movies, the missing at random hypothesis imputes missing ratings as the result of a random process that selects the items that are rated or not. This assumption might not be valid [9], causing the system to underperform. Indeed it is possible that an item, let's say a movie, is watched and then rated for a variety of reasons:

- the movie is popular (and the user often watches popular movies; although he might not necessarily like them),
- the movie is perceived by the user as similar to others seen in the past,
- the user thinks (based on his knowledge) that he might like the movie and therefore decide to watch it,
- the movie was recommended to the user (perhaps by a competitor), etc.

Therefore, instead of a simple probabilistic model, one could consider a richer model accounting for a mixture of all such different "user protocols" and the associated uncertainties in terms of subjective probabilities (allowing to model the interplay between the user habits, the popularity of movies, the beliefs of the user about which movies he might like, etc). Of course, learning such a probabilistic model would be challenging. We believe that the design of recommender systems dealing with such "protocol uncertainty" is an important research direction.

References

1. Grünwald, P., Halpern, J.Y.: Updating probabilities. J. Artif. Intell. Res. **19**, 243–278 (2003). https://doi.org/10.1613/jair.1164, https://doi.org/10.1613/jair.1164
2. Halpern, J.Y.: Reasoning about uncertainty. MIT Press (2005)
3. Jaeger, M.: On testing the missing at random assumption. In: Machine Learning: ECML 2006, 17th European Conference on Machine Learning, Berlin, Germany, September 18-22, 2006, Proceedings. pp. 671–678 (2006)
4. Jensen, F.V., Nielsen, T.D.: Bayesian Networks and Decision Graphs. Springer Publishing Company, Incorporated, 2nd edn. (2007)
5. Kooi, B.P.: Probabilistic dynamic epistemic logic. Journal of Logic, Language and Information **12**(4), 381–408 (2003)
6. Lucas, S.K., Rosenhouse, J.: Optimal strategies for the progressive Monty Hall problem. The Mathematical Gazette **93**(528), 410–419 (2009). https://doi.org/10.1017/S0025557200185158
7. Lucas, S.K., Rosenhouse, J., Schepler, A.: The Monty Hall problem, reconsidered. Mathematics Magazine **82**(5), 332–342 (2009), http://www.jstor.org/stable/27765931
8. MacKay, D.J.C.: Information Theory, Inference, and Learning Algorithms. Cambridge University Press (2003)
9. Marlin, B.M., Zemel, R.S.: Collaborative prediction and ranking with non-random missing data. In: Proceedings of the 2009 ACM Conference on Recommender Systems, RecSys 2009, New York, NY, USA, October 23-25, 2009. pp. 5–12. ACM (2009)
10. Mueser, P., Granberg, D., Mueser, K., Nickerson, R.: The Monty Hall dilemma revisited: Understanding the interaction of problem definition and decision making. University of Missouri Working Paper (02 1999)
11. Rosenhouse, J.: The Monty Hall Problem: The Remarkable Story of Math's Most Contentious Brainteaser. Oxford University Press (2009)
12. vos Savant, M.: Ask marilyn. Parade p. 15 (September 1990)
13. vos Savant, M.: Ask marilyn. Parade p. 25 (December 1990)
14. Selvin, S., Bloxham, M., Khuri, A.I., Moore, M., Coleman, R., Bryce, G.R., Hagans, J.A., Chalmers, T.C., Maxwell, E.A., Smith, G.N.: Letters to the editor. The American Statistician **29**(1), pp. 67–71 (1975), http://www.jstor.org/stable/2683689
15. Stibel, J., Dror, I., Ben-Zeev, A.: The collapsing choice theory: Dissociating choice and judgment in decision making. Theory and Decision **66**, 149–179 (02 2009). https://doi.org/10.1007/s11238-007-9094-7
16. Viappiani, P., Boutilier, C.: On the equivalence of optimal recommendation sets and myopically optimal query sets. Artificial Intelligence Journal **286**, 103328 (2020)