One game show, two boys, two aces, three prisoners - what’s an AI to do?

Eric Neufeld* and Sonje Finnestad
Department of Computer Science, University of Saskatchewan, Saskatoon, SK CANADA
*eric.neufeld@usask.ca

Abstract
We review a quartet of widely discussed probability puzzles – Monty Hall, the three prisoners, the two boys, and the two aces. Pearl explains why the Monty Hall problem is counterintuitive using a causal diagram. Glenn Shafer uses the puzzle of the two aces to justify reintroducing to probability theory protocols that specify how the information we condition on is obtained. Pearl, in one treatment of the three prisoners, adds to his representation random variables that distinguish actual events and observations. The puzzle of the two boys took a perplexing twist in 2010. We show the puzzles have similar features, and each can be made to give different answers to simple queries corresponding to different presentations of the word problem. We offer a unified treatment that explains this phenomenon in strictly technical terms, as opposed to cognitive or epistemic.

Introduction
In the Monty Hall Puzzle, there are three doors. Behind one is a brand-new car, and behind the other two are goats. After the contestant selects one door at random, the host opens one of the other two, revealing a goat. The host gives the contestant the opportunity to “switch or stay”. What should the contestant do?

As Pearl and Mackenzie recently (2018) document, this generated an unexpected controversy when it appeared in a puzzle column by Marilyn vos Savant (1990), who argued that switching doors doubles the contestant’s chances of winning. She illustrated her solution with a small table like that shown in Table 1.

<table>
<thead>
<tr>
<th>Door 1</th>
<th>Door 2</th>
<th>Door 3</th>
<th>Door Opened</th>
<th>Outcome if You Switch</th>
<th>Outcome if You Stay</th>
</tr>
</thead>
<tbody>
<tr>
<td>auto</td>
<td>goat</td>
<td>goat</td>
<td>2 or 3</td>
<td>lose</td>
<td>win</td>
</tr>
<tr>
<td>goat</td>
<td>auto</td>
<td>goat</td>
<td>3</td>
<td>win</td>
<td>lose</td>
</tr>
<tr>
<td>goat</td>
<td>goat</td>
<td>auto</td>
<td>2</td>
<td>win</td>
<td>lose</td>
</tr>
</tbody>
</table>

Table 1. Outcomes for switching and staying

Vos Savant’s solution was widely and hotly disputed (Burns & Wieth, 2004; vos Savant, 1997), most arguing that once the host opened a door, the prize was equally likely to be behind the original door and the unopened door.

The AI community remains interested in this puzzle, partly because it raises questions about human cognitive processes and intuitions. Pearl and MacKenzie explain why this problem rubs intuition the wrong way using the causal graph below. Those familiar with causal graphs (a.k.a Bayes Nets) understand that variables YourDoor and LocationOfCar are probabilistically independent by construction, but also independent common causes of DoorOpened.

Your door  Location of Car

Door Opened

It is also well known in Bayes net lore that because of the colliders (head-to-head arrows) at DoorOpened, observing DoorOpened creates a probabilistic association between two independent causes resulting in a clash of intuitions.

Falk (2011) looks at the two-boys puzzle from the perspective of cognitive science. It may be posed as follows: in a world where all families have exactly two children, Ms. Jones has at least one boy. What is the probability Ms. Jones’s other child is a boy (Gardner, 1961)?

Many argue the answer is 1/2, because the gender of an individual is independent of the gender of any other. Yet others (Bellos, 2019; Gardner, 1961; vos Savant in Stansfield & Carlton, 2009) argue the answer is 1/3: Bellos and vos Savant also conducted surveys to support their arguments. Letting M indicate male and F female gives us represent four equally likely two-child families: FF, FM, MF, MM, (letter order represents birth order). Knowing Ms. Jones has a boy rules out FF, leaving three equally likely possible families, just one of which has a second boy. Bellos (2019) offers several intriguing variations on this latter methodology. If Ms. Jones’s eldest is a boy, the probability of two boys becomes ½. But if she has a boy born on a Tuesday, the probability of two boys is 13/27 – almost, but not quite 1/2. Falk (2011) pursues this in detail, giving a remarkable formula showing that the more improbable the feature observed (e.g., the second of birth), the closer the probability of two boys gets to 1/2, writing “individuality is characterised mathematically by an extremely narrow specification whose probability is infinitesimal. Such a unique specification of a boy turns out to be equivalent to observing him in person… Learning [that the boy was born on a Tuesday] lent some uniqueness to that son.” For those in artificial intelligence this raises the question whether automated solutions to word problems ensconced in real-world settings need cognitive or epistemic to deducing such features from natural language descriptions. Russell (2019), for example, suggests that a prerequisite for superintelligent machines is the ability to
learn technical material quickly by reading books. If so, does a computer need to pass the Turing Test first (Neufeld and Finnestad, 2020a, 2020b) to solve questions like this? It seems counterintuitive that the more irrelevant observations we make, the closer we get the intuitively correct answer.

A formal look at the two boys puzzle

Bar-Hillel and Falk (1982) first explored this puzzle using a sample space of two-child family kinds \{FF, FM, MF, MM\}, each with probability 1/4.1

To obtain a solution of 1/3 we imagine a knowledge-seeker who has “come to know” the event atLeastOneBoy = \{FM, MF, MM\}, logically and probabilistically equivalent to \~FF. By construction, the probability of FF is 1/4, so \(p(\~FF) = 3/4\).

The probability of two boys given at least one boy is

\[
p(MM|\~FF) = p(FF|MM) + p(MF|MM) p(MF) + p(MF|MM) p(MM)\]

\[
= 0 + 1/2 * 1/4 + 1/2 * 1/4 + 1/4 = 1/2.
\]

Next, the knowledge-seeker computes the probability that the remaining child in the drawn urn is a male given \(m_i\), which is equivalent to the urn being MM, so our target probability is:

\[
p(MM|m_i) = p(m_i|MM) p(MM) / p(m_i)\]

\[
= 1 * 1/4 / (1/2) = 1/2.
\]

different a value for the vague linguistic expression “the probability of two boys given at least one boy”.2

In Solution 2, birth order is immaterial, as in Solution 1 – it lets us create four equiprobable sample space elements.

It’s not wrong, however the next section considers a domain where Solution 1 may be more natural.

1 Questions of gender and sex are complex and highly contested. However, Falk further states that the model is close to certain empirical distributions, and also has pedagogical merit.

2 Gelman (2010) points out that, empirically, days of birth are not equally likely. Regardless, a natural assumption would be that the two variables are independent.

3 Request the full paper for a proof.
**The two aces**

Shafer’s (1985) mentions this problem in making a case for the reintroduction of protocols to probability. A four-card deck consists of the ace and deuce of hearts and the ace and deuce of spades. The dealer shuffles, then deals two cards to a colleague.

The dealer asks, “do you have an ace?” The colleague replies, “yes.” The dealer’s belief that the colleague has two aces changes from 1/6 (all hands are equally likely) to 1/5 (all hands excluding the two-deuce hand).

If the dealer initially asks, “do you have the ace of hearts?” and the colleague answers, “yes”, the colleague holds one of only three hands, and the dealer’s belief that the colleague holds two aces changes 1/3.

But consider Falk’s reasoning about the Tuesday boy – that the boy first picked must be born on some day – in this setting. If the colleague answers “Yes” to the first question, the colleague must be holding either the ace of hearts or the ace of spades, that is, some ace, again suggesting a “proof by cases” the correct answer is 1/3.

To make our two solution approaches realistic in this setting, we use two different physical models. For Solution 1, the dealer prepares six slips of paper, each displaying one of six possible hands. Instead of two cards, the dealer gives the colleague one slip of paper – after a good shuffle, of course. If the colleague replies “yes” when asked whether the colleague holds an ace, five possibilities clearly remain and the probability of at least one ace is 5/6 at this point. This “coming to know” at least one ace parallels Solution 1 and the course. If one of six possible hands.

1

Setting by cases” the correct answer is 1/3.

1

Setting.

That the boy first picked m

1

Setting.

Deuce of spades.

The reintroduction o

Shafer

For

Solution 2

Shafer illustrates his puzzle with “yes/no” questions but note that the other problems share the feature that no information is obtained by asking “yes/no” questions. Prisoner A’s circumstances force the contrivance of an indirect question that neither confirms or denies that the prisoner A will be set free. Our introduction of an initial draw of ~FF to the two boys puzzle is also a contrivance that provides a way for the knowledge-seeker to learn ~FF and no more.

Shafer then gives the example of the dealer asking if the colleague holds an ace, to which the colleague replies “yes”, and adds with a smile, “in fact, I hold the ace of hearts.”

**Prisoner A imagines that the guard might be convinced that naming B or C as one who will not be freed doesn’t violate the instructions, as A would not be able to deduce A’s own fate with certainty. However, the additional information may be useful.

Letting $F_A$ mean “Prisoner A will be freed” and $\neg F_B$ mean “Prisoner B will not.” Then

$$p(F_A | \neg F_B) = p(\neg F_B | F_A) \frac{p(F_A)}{p(\neg F_B)} = 1^{*} \frac{2}{3}/2 = \frac{1}{3},$$

momentarily cheering the prisoner. But contemplating further, A realizes that A’s probability of freedom, should the guard reveal C will not go free, is also 1/2. Just by thinking about the problem, A finds a “proof by cases” that A’s probability of freedom is now 1/2 instead of 1/3.

The technical problem is that $p(\neg F_B | F_A)$ shouldn’t enter this calculation. If $F_A$, the guard may choose either $\neg F_B$ or $\neg F_C$, but $p(\neg F_B | F_C) = 1$. Pearl (1988) instead conditions on actual observations and not their implications, using a distinct variable that reports an observation; e.g., $\neg F_B$ is the new proposition that the guard reports that $B$ will not go free, and isn’t implied by $F_A$. The equation now becomes

$$p(F_A | \neg F_B) = p(\neg F_B | F_A) \frac{p(F_A)}{p(\neg F_B)} = 1/2 \times (1/3) / (1/2) = 1/3.$$
Suppose the knowledge-seeker in the two boys puzzle asks “is there at least one boy?” and a mischievous oracle (corresponding to the smiling colleague) answers “Yes”, then adds “as a matter of fact, the eldest is a boy.” what then is the correct answer?

Shafer argues that this shows the need for protocols that consider all possible answers to all possible questions, including volunteered information, tones, and tells.

Space does not permit lengthier discussion, but we remark it links Shafer’s protocols with Pearl and MacKenzie’s (2018) refer to as the “data-generating process.”

Conclusions and Future Work

The present work began with a rhetorical question: how should an AI resolve ambiguities inherent in word problems, in particular about probability. It arose from a look at the work of Falk (2011) on the Tuesday boy problem, and attempts to find a purely technical solution that did not need to make understand cognitive or epistemic matters. Although we produced a technical solution to the “Tuesday boy” problem that though our solution to the Tuesday boy problem that did not rely on cognitive or epistemic notions of “individuality”, solutions may vary depending on minor details of setting and the “obvious” answer may rely on an understanding of the different scenarios. Space didn’t allow us to expand on the issue of probabilistic “proof by cases”.

Despite the artificiality of these puzzles, the solution to real problems are inevitably far more complex, and depend on asking questions that guide agents toward good answers, appropriately evaluating those answers, or understanding how we came to know certain information (Kyburg, 1984).

Acknowledgments

Thanks to the many readers of earlier version of this paper, and to the careful and insightful comments of the reviewers.

References


