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?

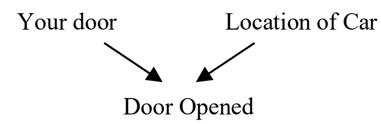
As Pearl and Mackenzie recently (2018) document, this generated an unexpected controversy when it appeared in a puzzle column by Marilyn vos Savant (1990), who argued that switching doors doubles the contestant's chances of winning. She illustrated her solution with a small table like that shown in Table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

Door 1	Door 2	Door 3	Door Opened	Outcome if You Switch	Outcome if You Stay
auto	goat	goat	2 or 3	lose	win
goat	auto	goat	3	win	lose
goat	goat	auto	2	win	lose

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*.



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 $p(\sim FF)$ is 3/4.

The probability of two boys given at least one boy is $p(MM|\sim FF)=p(\sim FF|MM)\,p(MM)/p(\sim FF)=1*1/4/(3/4)= 1/3$, by Bayes' Rule. Bar-Hillel and Falk (1982) suggest once the outcome *FF* is eliminated from the sample space, and the remaining outcomes are equiprobable at 1/3. Falk relates the convincing example of meeting parents of members of a boy scout troop as "coming to know ~FF'. This is Solution 1.

Solution 2 uses a balls and urns metaphor to obtain an answer of 1/2. Consider *FF, FM, MF, MM* as urns, each containing one family kind, the children represented by balls. First the knowledge-seeker draws an urn (family), then draws a ball (child). Let m_1 denote that this first child drawn is a boy. The prior of m_1 is

$$p(m_1) = p(m_1|FF)p(FF) + p(m_1|FM)p(FM) + p(m_1|MF)p(MF) + p(m_1|MM)\,p(MM)$$
$$= 0 + 1/2 * 1/4 + 1/2 * 1/4 + 1 * 1/4 = 1/2.$$

Next, the knowledge-seeker computes the probability that the remaining child in the drawn urn is a male given m_1, which is equivalent to the urn being *MM*, so our target probability is:

$$p(MM|m_1) = p(m_1|MM)\,p(MM) / p(m_1)$$
$$= 1 * 1/4 / (1/2), = ½$$

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

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

Birth order plays a role if we observe the eldest child's gender. Using the method of Solution 1, this amounts to observing the event {*MF, MM*} and the probability of two boys is clearly 1/2. Here birth order is important.

The method of Solution 2 also yields 1/2, but by a different route. First, the knowledge-seeker draws a family, then observes the eldest in that family. *If* the eldest is a boy, the knowledge-seeker computes the probability of em_1, that the remaining child is a boy:

$$p(em_1) = p(em_1|FF)p(FF) + p(em_1|FM)p(FM) + p(em_1|MF)p(MF) + p(em_1|MM)\,p(MM)$$
$$= 0 + 0 + 1/2 * 1/4 + 1/2 * 1/4 = 1/4.$$

Then, using Bayes' rule,

$$p(MM|em_1) = p(em_1|MM)\,p(MM) / p(em_1)$$
$$= 1/2 * 1/4 / (1/4) = 1/2.$$

Both approaches give the same answer. Now consider the problem of a boy born on a Tuesday.

Falk (2011) presents the Tuesday boy as a curious extension of Solution 1. Assuming a child is equally likely to be born on any day[2] independently of gender, a straightforward but tedious approach is to create a sample space with 196 elements, each representing one of all the possible combinations of the first and second child's gender and day of birth. If we collapse all females into a single kind, and all boys not born on a Tuesday to a single kind, we end up with the following count of family kinds:

FM_T	7	M_TM_T	1	$M_{\sim T}M_T$	6
M_TF	7	$M_TM_{\sim T}$	6	*REST*	169

FM_T indicates a family whose eldest is *F* and whose youngest is a Tuesday male; $m_{\sim T}$ is "a male not born on a Tuesday". (This simple enumeration gives an early warning that the answer cannot be ½ as the number of families with a Tuesday boy is odd.)

Using the method of Solution 1, the knowledge-seeker observes *boyBornOnATuesday*, the subset containing the first five family kinds in the sample space. Summing up, this event has probability 27/196. We then compute the probability of "two boys" $\{M_TM_T, M_TM_{\sim T}, M_{\sim T}M_T\}$, which is 13/196, which yields 13/27 as the probability of two boys, given (at least) one boy born on a Tuesday.

The result is troubling because, as Falk says, we could derive 13/27 for each day of the week, and then use "proof by cases" to argue that if one child is a boy, it must be born on *some* day, and thus the other child is a boy with probability 13/27 rather than 1/3 *or* 1/2 – even more troubling. Intuitions suggests day of birth is irrelevant to gender. The sample space was constructed based on that assumption. Falk then shows that as increasingly rare features (hour or second of birth) are chosen, the probability tends to 1/2.

Now consider Solution 2. The knowledge-seeker draws a family, then draws a child that happens to be a Tuesday boy. If the reader follows the method of Solution 2 meticulously. they will find the probability of two boys is *exactly* 1/2.[3] Not only is this answer intuitive, it also is consistent with day of birth and gender being independent'

Butt what of the surveys of Bellos and vos Savant mentioned earlier that seem to empirically support Solution 1? The distribution of the surveys was constructed in such a way that it reflects the design of the sample space. We believe our Solution 2 derivations give evidence that Solution 1 is not well-suited to *this particular domain*. 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 becomes 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 in the two boys puzzle. The probability of two aces becomes 1/5. (Notice that if you multiply these two numbers together, the result is 1/6, which provides a check on the reasoning.) A similar argument can be made for the scenario where the dealer asks "Do you have the ace of hearts?"

For Solution 2, cards are dealt one at a time, face down on the table. The dealer asks the colleague to draw one card, and asks if it is an ace.

Like Solution 2 to the two boys puzzle, Solution 2 involves two draws: first *of* the hand, then *from* the hand. If the answer to the question is 'yes', the probability of two aces becomes 1/3.

In this domain the approach of Solution 1 may be more intuitively satisfying.

A completely new problem: Shafer (1985) asks, suppose in Solution 1, the dealer asks if colleague has an ace, and the colleague replies, "Yes", then adds, "in fact, I have the ace of hearts," *while smiling*, showing another way probabilities changes when unasked-for information is received'

The three prisoners

Three prisoners A, B, and C discover a monarch will grant clemency to exactly one of them. The probability of clemency is 1/3 for each. The prison guard knows who will be freed but is under strict instructions not to give any prisoner information that reveals *that* prisoner's fate.

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 $\sim F_B$ mean "Prisoner B will not." Then

$$p(F_A|\sim F_B)=p(\sim F_B|F_A)\,p(\mathrm{F_A})/p(\sim F_B)=1*(1/3)/(2/3) = ½,$$

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 $p(\sim F_B|F_A)$ shouldn't enter this calculation. If F_A, the guard may choose either $\sim F_B$ or $\sim F_C$, but $p(\sim F_B|F_A)$=1. Pearl (1988) instead conditions on actual observations and not their implications, using a distinct variable that reports an observation; e.g., $\sim F_B'$ 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

$$p(F_A|\sim F_B')=p(\sim F_B'|F_A)p(\mathrm{F_A})/p(\sim F_B')=1/2*(1/3)/(1/2)= 1/3.$$

Hearing the guard's answer is analogous to seeing the door open in the Monty Hall problem and 2/3 of the time, the prisoner not mentioned by the guard (and excluding A) is the one who will go free.

This distinction between observations and actual events makes an appearance in all the preceding puzzles.

Revisiting the two boys

A commonality between the two boys and the three prisoners is in the way information is received. We have already remarked it is difficult to imagine ways a real-life knowledge-seeker discovers only $\sim FF$ without discovering something about one of the children: perhaps we could label the *FF* urn with its name, and label the other three $\sim FF$. If an urn labelled $\sim FF$ is drawn and the probability of *MM* is 1/3. But if the knowledge-seeker then draws from the urn and observes a boy, the probability of *MM* is now 1/2 because the draw of a child *might* produce a girl, (just as the guard might give two different answers) yielding a very different distribution: *MM* would be 0, *FM* and *MF* would both be 1/2).

This unifies the two boys, the three prisoners and Monty Hall.

Revisiting the two aces

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 $\sim FF$ to the two boys puzzle is also a contrivance that provides a way for the knowledge-seeker to learn $\sim FF$ 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."

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

Bar-Hillel, M., & Falk, R. (1982). Some Teasers Concerning Conditional Probabilities. In *Cognition*, *11*(2), 109-122.

Bellos, A. (2019, November 18). Did you solve it? The two child problem. *The Guardian*, Retrieved from www.theguardian.com.

Burns, B. D., & Wieth, M. (2004). The collider principle in causal reasoning: why the Monty Hall dilemma is so hard. In *Journal of Experimental Psychology: General*, *133*(3), 434.*Proceedings of the 2nd International Conference*, San Mateo, pp. 441-452

Falk, R. (1992). A closer look at the probabilities of the notorious three prisoners. In *Cognition, 43,* 197–223.

Falk, R. (2011). When truisms clash: Coping with a counterintuitive problem concerning the notorious two-child family. In *Thinking & Reasoning*, *17*(4), 353-366.

Gardner, M. (1961). *The Second Scientific American Book of Mathematical Puzzles and Diversions*. Simon & Schuster.

Gelman, Andrew. (2010, May 27). Hype about conditional probability puzzles [Blog post]. Retrieved from https://statmodeling.stat.columbia.edu/2010/05/27/hype_about_cond/.

Kallenberg, O. (1974). A note on the asymptotic equivalence of sampling with and without replacement. In *The Annals of Statistics*, 819-821.

Kyburg, H. E., Jr. (1984) *Theory and Measurement* Cambridge University Press.

Neufeld, E. & Finnestad, S. (2020a) Imitation Game: Threshold or Watershed? *Minds and Machines* 30(4) 637-657

Neufeld, E. & Finnestad, S. (2020b) In defense of the Turing Test, *Artificial Intelligence & Society* 35 819-827.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc. San Francisco, CA

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. New York: Cambridge University Press.

Pearl, J. (2014). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Elsevier.

Pearl, J., & Mackenzie, D. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic books.

Russell, S. (2019). *Human Compatible: Artificial Intelligence and the Problem of Control*. Penguin.

Shafer, G. (1985). Conditional Probability. In *International Statistical Review / Revue Internationale De Statistique, 53*(3), 261-275.

Stansfield, William D, Carlton, Matthew A. (February 2009). The Most Widely Publicized Gender Problem in Human Genetics. In *Human Biology*, 81 (1), 3–11.

vos Savant, M. (1990, September 9). Ask Marilyn. In *Parade Magazine,* 15.

vos Savant, M. (1997). *The power of logical thinking*. New York: St Martin’s Press.