

Boy or Girl paradox

From Wikipedia, the free encyclopedia

The **Boy or Girl paradox** surrounds a set of questions in probability theory which are also known as *The Two Child Problem*,^[1] *Mr. Smith's Children*^[2] and the *Mrs. Smith Problem*. The initial formulation of the question dates back to at least 1959, when Martin Gardner published one of the earliest variants of the paradox in *Scientific American*. Titled *The Two Children Problem*, he phrased the paradox as follows:

- Mr. Jones has two children. The older child is a girl. What is the probability that both children are girls?
- Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?

Gardner initially gave the answers $\frac{1}{2}$ and $\frac{1}{3}$, respectively, but later acknowledged that the second question was ambiguous.^[3] Its answer could be $\frac{1}{2}$, depending on how you found out that one child was a boy. The ambiguity, depending on the exact wording and possible assumptions, was confirmed by Bar-Hillel and Falk,^[4] and Nickerson.^[5]

Other variants of this question, with varying degrees of ambiguity, have been popularized by [Ask Marilyn](#) in *Parade Magazine*,^[6] [John Tierney](#) of *The New York Times*,^[7] and Leonard Mlodinow in *Drunkard's Walk*.^[8] One scientific study showed that when identical information was conveyed, but with different partially ambiguous wordings that emphasized different points, that the percentage of [MBA](#) students who answered $\frac{1}{2}$ changed from 85% to 39%.^[2]

The paradox has frequently stimulated a great deal of controversy.^[5] Many people argued strongly for both sides with a great deal of confidence, sometimes showing disdain for those who took the opposing view. The paradox stems from whether the problem setup is similar for the two questions.^{[2][8]} The intuitive answer is $\frac{1}{2}$.^[2] This answer is intuitive if the question leads the reader to believe that there are two equally likely possibilities for the sex of the second child (i.e., boy and girl),^{[2][9]} and that the probability of these outcomes is absolute, not [conditional](#).^[10]

Contents

- [1 Common assumptions](#)
- [2 First question](#)
- [3 Second question](#)
- [4 Analysis of the ambiguity](#)
- [5 Bayesian analysis](#)
- [6 Martingale analysis](#)
- [7 Variants of the question](#)
 - [7.1 Information about the child](#)
- [8 Psychological investigation](#)
- [9 See also](#)
- [10 References](#)
- [11 External links](#)

Common assumptions

The two possible answers share a number of assumptions. First, it is assumed that the space of all possible events can be easily enumerated, providing an **extensional definition** of outcomes: {BB, BG, GB, GG}.^[11] This notation indicates that there are four possible combinations of children, labeling boys B and girls G, and using the first letter to represent the older child. Second, it is assumed that these outcomes are equally probable.^[11] This implies the following **model**, a **Bernoulli process** with $p = \frac{1}{2}$:

1. Each child is either male or female.
2. Each child has the same chance of being male as of being female.
3. The sex of each child is independent of the sex of the other.

The mathematical outcome would be the same if it were phrased in terms of a **coin toss**.

First question

- Mr. Jones has two children. The older child is a girl. What is the probability that both children are girls?

Under the forementioned assumptions, in this problem, a random family is selected. In this sample space, there are four *equally probable* events:

Older child	Younger child
Girl	Girl
Girl	Boy
Boy	Girl
Boy	Boy

Only two of these possible events meet the criteria specified in the question (i.e., GG, GB). Since both of the two possibilities in the new sample space {GG, GB} are equally likely, and only one of the two, GG, includes two girls, *the probability that the younger child is also a girl is $\frac{1}{2}$* .

Second question

- Mr. Smith has two children. At least one of them is a boy. What is the probability that both children are boys?

This question is identical to question one, except that instead of specifying that the older child is a boy, it is specified that at least one of them is a boy. In response to reader criticism of the question posed in 1959, Gardner agreed that a precise formulation of the question is critical to getting different answers for question 1 and 2. Specifically, Gardner argued that a "failure to specify the randomizing procedure" could lead readers to interpret the question in two distinct ways:

- From all families with two children, at least one of whom is a boy, a family is chosen at random. This would yield the answer of $\frac{1}{3}$.
- From all families with two children, one child is selected at random, and the sex of that child is specified to be a boy. This would yield an answer of $\frac{1}{2}$.^{[4][5]}

Grinstead and Snell argue that the question is ambiguous in much the same way Gardner did.^[12]

For example, if you see the children in the garden, you may see a boy. The other child may be hidden behind a tree. In this case, the statement is equivalent to the second (the child that you can see is a boy). The first statement does not match as one case is one boy, one girl. Then the girl may be visible. (The first statement says that it can be either.)

While it is certainly true that every possible Mr. Smith has at least one boy (i.e., the condition is necessary) it is not clear that every Mr. Smith with at least one boy is intended. That is, the problem statement does not say that having a boy is a sufficient condition for Mr. Smith to be identified as having a boy this way.

Commenting on Gardner's version of the problem, Bar-Hillel and Falk^[4] note that "Mr. Smith, unlike the reader, is presumably aware of the sex of both of his children when making this statement", i.e. that 'I have two children and at least one of them is a boy.' If it is further assumed that Mr Smith would report this fact if it were true then the correct answer is $\frac{1}{3}$ as Gardner intended.

Analysis of the ambiguity

If it is assumed that this information was obtained by looking at both children to see if there is at least one boy, the condition is both necessary and sufficient. Three of the four equally probable events for a two-child family in the sample space above meet the condition, as in this table:

Older child	Younger child
Girl	Girl
Girl	Boy
Boy	Girl
Boy	Boy

Thus, if it is assumed that both children were considered while looking for a boy, the answer to question 2 is $\frac{1}{3}$. However, if the family was first selected and *then* a random, true statement was made about the gender of one child in that family, whether or not both were considered, the correct way to calculate the conditional probability is not to count all of the cases that include a child with that gender. Instead, one must consider only the probabilities where the statement will be made in each case.^[12] So, if **ALOB** represents the event where the statement is "at least one boy", and **ALOG** represents the event where the statement is "at least one girl", then this table describes the sample space:

Older child	Younger child	P(this family)	P(ALOB given this family)	P(ALOG given this family)	P(ALOB and this family)	P(ALOG and this family)
Girl	Girl	$\frac{1}{4}$	0	1	0	$\frac{1}{4}$
Girl	Boy	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$
Boy	Girl	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{8}$	$\frac{1}{8}$
Boy	Boy	$\frac{1}{4}$	1	0	$\frac{1}{4}$	0

So, if you are told that at least one is a boy when the fact is chosen randomly, the probability that both are boys is

$$\frac{P(\text{ALOB and BB})}{P(\text{ALOB})} = \frac{\frac{1}{4}}{0 + \frac{1}{8} + \frac{1}{8} + \frac{1}{4}} = \frac{1}{2}.$$

The paradox occurs when it is not known how the statement "at least one is a boy" was generated. Either answer could be correct, based on what is assumed.^[13]

However, the " $\frac{1}{3}$ " answer is obtained only by assuming $P(\text{ALOB}|\text{BG}) = P(\text{ALOB}|\text{GB}) = 1$, which implies $P(\text{ALOG}|\text{BG}) = P(\text{ALOG}|\text{GB}) = 0$, that is, the other child's sex is never mentioned although it is present. As Marks and Smith say, "This extreme assumption is never included in the presentation of the two-child problem, however, and is surely not what people have in mind when they present it."^[13]

Bayesian analysis

Following classical probability arguments, we consider a large urn containing two children. We assume equal probability that either is a boy or a girl. The three discernible cases are thus: 1. both are girls (GG) — with probability $P(\text{GG}) = \frac{1}{4}$, 2. both are boys (BB) — with probability of $P(\text{BB}) = \frac{1}{4}$, and 3. one of each (G·B) — with probability of $P(\text{G·B}) = \frac{1}{2}$. These are the *prior* probabilities.

Now we add the additional assumption that "at least one is a boy" = B. Using [Bayes' Theorem](#), we find

$$P(\text{BB} | \text{B}) = P(\text{B} | \text{BB}) \times \frac{P(\text{BB})}{P(\text{B})} = 1 \times \frac{\left(\frac{1}{4}\right)}{\left(\frac{3}{4}\right)} = \frac{1}{3}.$$

where $P(\text{A}|\text{B})$ means "probability of A given B". $P(\text{B}|\text{BB})$ = probability of at least one boy given both are boys = 1. $P(\text{BB})$ = probability of both boys = $\frac{1}{4}$ from the prior distribution. $P(\text{B})$ = probability of at least one being a boy, which includes cases BB and G·B = $\frac{1}{4} + \frac{1}{2} = \frac{3}{4}$.

Note that, although the natural assumption seems to be a probability of $\frac{1}{2}$, so the derived value of $\frac{1}{3}$ seems low, the actual "normal" value for $P(\text{BB})$ is $\frac{1}{4}$, so the $\frac{1}{3}$ is actually a bit *higher*.

The paradox arises because the second assumption is somewhat artificial, and when describing the problem in an actual setting things get a bit sticky. Just how do we know that "at least" one is a boy? One description of the problem states that we look into a window, see only one child and it is a boy. This sounds like the same assumption. However, this one is equivalent to "sampling" the distribution (i.e. removing one child from the urn, ascertaining that it is a boy, then replacing). Let's call the statement "the sample is a boy" proposition "b". Now we have:

$$P(\text{BB} | \text{b}) = P(\text{b} | \text{BB}) \times \frac{P(\text{BB})}{P(\text{b})} = 1 \times \frac{\left(\frac{1}{4}\right)}{\left(\frac{1}{2}\right)} = \frac{1}{2}.$$

The difference here is the $P(\text{b})$, which is just the probability of drawing a boy from all possible cases (i.e. without the "at least"), which is clearly $\frac{1}{2}$.

The Bayesian analysis generalizes easily to the case in which we relax the 50:50 population assumption. If we have no information about the populations then we assume a "flat prior", i.e. $P(\text{GG}) = P(\text{BB}) = P(\text{G·B}) = \frac{1}{3}$. In this case the "at least" assumption produces the result $P(\text{BB}|\text{B}) = \frac{1}{2}$, and the sampling assumption produces $P(\text{BB}|\text{b}) = \frac{2}{3}$, a result also derivable from the [Rule of Succession](#).

Martingale analysis

Suppose you had wagered that Mr Smith had two boys, and received fair odds. You paid \$1 and you will receive \$4 if he has two boys. We think of your wager as investment that will increase in value as good news arrives. What evidence would make you happier about your investment? Learning that at least one child out of two is a boy, or learning that at least one child out of one is a boy?

The latter is *a priori* less likely, and therefore better news. That is why the two answers cannot be the same.

Now for the numbers. If we bet on one child and win, the value of your investment has doubled. It must double again to get to \$4, so the odds are 1 in 2.

On the other hand if we learn that at least one of two children is a boy, our investment increases as if we had wagered on this question. Our \$1 is now worth $\$1\frac{1}{3}$. To get to \$4 we still have to increase our wealth threefold. So the answer is 1 in 3.

Variants of the question

Following the popularization of the paradox by Gardner it has been presented and discussed in various forms. The first variant presented by Bar-Hillel & Falk^[4] is worded as follows:

- Mr. Smith is the father of two. We meet him walking along the street with a young boy whom he proudly introduces as his son. What is the probability that Mr. Smith's other child is also a boy?

Bar-Hillel & Falk use this variant to highlight the importance of considering the underlying assumptions. The intuitive answer is $\frac{1}{2}$ and, when making the most natural assumptions, this is correct. However, someone may argue that "...before Mr. Smith identifies the boy as his son, we know only that he is either the father of two boys, BB, or of two girls, GG, or of one of each in either birth order, i.e., BG or GB. Assuming again independence and equiprobability, we begin with a probability of $\frac{1}{4}$ that Smith is the father of two boys. Discovering that he has at least one boy rules out the event GG. Since the remaining three events were equiprobable, we obtain a probability of $\frac{1}{3}$ for BB."^[4]

The natural assumption is that Mr. Smith selected the child companion at random. If so, as combination BB has twice the probability of either BG or GB of having resulted in the boy walking companion (and combination GG has zero probability, ruling it out), the union of events BG and GB becomes equiprobable with event BB, and so the chance that the other child is also a boy is $\frac{1}{2}$. Bar-Hillel & Falk, however, suggest an alternative scenario. They imagine a culture in which boys are invariably chosen over girls as walking companions. In this case, the combinations of BB, BG and GB are assumed *equally* likely to have resulted in the boy walking companion, and thus the probability that the other child is also a boy is $\frac{1}{3}$.

In 1991, Marilyn vos Savant responded to a reader who asked her to answer a variant of the Boy or Girl paradox that included beagles.^[6] In 1996, she published the question again in a different form. The 1991 and 1996 questions, respectively were phrased:

- A shopkeeper says she has two new baby beagles to show you, but she doesn't know whether they're male, female, or a pair. You tell her that you want only a male, and she telephones the fellow who's giving them a bath. "Is at least one a male?" she asks him. "Yes!" she informs you with a smile. What is the probability that the other one is a male?
- Say that a woman and a man (who are unrelated) each has two children. We know that at least one of the woman's children is a boy and that the man's oldest child is a boy. Can you explain why the chances that the woman has two boys do not equal the chances that the man has two boys?

With regard to the second formulation Vos Savant gave the classic answer that the chances that the woman has two boys are about $\frac{1}{3}$ whereas the chances that the man has two boys are about $\frac{1}{2}$. In response to reader response that questioned her analysis vos Savant conducted a survey of readers with exactly two children, at

least one of which is a boy. Of 17,946 responses, 35.9% reported two boys.^[11]

Vos Savant's articles were discussed by Carlton and Stansfield^[11] in a 2005 article in *The American Statistician*. The authors do not discuss the possible ambiguity in the question and conclude that her answer is correct from a mathematical perspective, given the assumptions that the likelihood of a child being a boy or girl is equal, and that the sex of the second child is independent of the first. With regard to her survey they say it "at least validates vos Savant's correct assertion that the "chances" posed in the original question, though similar-sounding, are different, and that the first probability is certainly nearer to 1 in 3 than to 1 in 2."

Carlton and Stansfield go on to discuss the common assumptions in the Boy or Girl paradox. They demonstrate that in reality male children are actually more likely than female children, and that the sex of the second child is not independent of the sex of the first. The authors conclude that, although the assumptions of the question run counter to observations, the paradox still has pedagogical value, since it "illustrates one of the more intriguing applications of conditional probability."^[11] Of course, the actual probability values do not matter; the purpose of the paradox is to demonstrate seemingly contradictory logic, not actual birth rates.

Information about the child

Suppose we were told not only that Mr. Smith has two children, and one of them is a boy, but also that the boy was born on a Tuesday: does this change the previous analyses? Again, the answer depends on how this information was presented - what kind of selection process produced this knowledge.

Following the tradition of the problem, suppose that in the population of two-child families, the sex of the two children is independent of one another, equally likely boy or girl, and that the birth date of each child is independent of the other child. The chance of being born on any given day of the week is $\frac{1}{7}$.

From Bayes' Theorem that the probability of two boys, given that one boy was born on a Tuesday is given by:

$$P(\mathbf{BB} \mid \mathbf{B}_T) = \frac{P(\mathbf{B}_T \mid \mathbf{BB}) \times P(\mathbf{BB})}{P(\mathbf{B}_T)}$$

Assume that the probability of being born on a Tuesday is ε ($\varepsilon = \frac{1}{7}$ will be set after arriving at the general solution). The first term in the numerator is therefore the probability of at least one boy born on Tuesday, given that the family has two boys, or $1 - (1 - \varepsilon)^2$ (one minus the probability that neither boy is born on Tuesday). The second term in the numerator is simply $\frac{1}{4}$, the probability of having two boys. The denominator is trickier; the required probability is that of having at least one boy on Tuesday over the entire sample space (two-child families). Therefore there are 4 cases to evaluate: BB, BG, GB, GG. Each of these occurs with probability $\frac{1}{4}$. $P(\mathbf{B}_T(\mathbf{GG}))$ is 0, there are no boys. $P(\mathbf{B}_T(\mathbf{BG}))$ and $P(\mathbf{B}_T(\mathbf{GB}))$ is ε , there is one and only one boy, thus he has ε chance of being born on Tuesday. $P(\mathbf{B}_T(\mathbf{BB}))$ is $\varepsilon + \varepsilon - \varepsilon^2$. It is the chance that one boy is born on Tuesday, plus the chance that the other boy is born on Tuesday, minus the chance that they both are (this term arises from the fact that $P(A \text{ or } B) = P(A) + P(B) - P(A)P(B)$, assuming that A and B are independent). Therefore, the full equation is:

$$P(\mathbf{BB} \mid \mathbf{B}_T) = \frac{(1 - (1 - \varepsilon)^2) \times \frac{1}{4}}{0 + \frac{1}{4}\varepsilon + \frac{1}{4}\varepsilon + \frac{1}{4}(\varepsilon + \varepsilon - \varepsilon^2)} = \frac{1 - (1 - \varepsilon)^2}{4\varepsilon - \varepsilon^2}$$

$$\text{For } \varepsilon > 0, \text{ this reduces to } P(\mathbf{BB} \mid \mathbf{B}_T) = \frac{2 - \varepsilon}{4 - \varepsilon}$$

If ε is now set to $\frac{1}{7}$, the probability becomes $\frac{13}{27}$, or about 0.48. In fact, as ε approaches 0, the total probability goes to $\frac{1}{2}$, which is the answer expected when one child is sampled (e.g. the oldest child is a boy) and is thus removed from the pool of possible children. In other words, as more and more details about the boy child are given (for instance: born on January 1), the chance that the other child is a girl approaches one half.

It seems that quite irrelevant information was introduced, yet the probability of the sex of the other child has changed dramatically from what it was before (the chance the other child was a girl was $\frac{2}{3}$, when it was not known that the boy was born on Tuesday).

However, is it really plausible that the family with at least one boy born on a Tuesday was produced by choosing just one of such families at random? It is much more easy to imagine the following scenario.

- We know Mr. Smith has two children. We knock at his door and a boy comes and answers the door. We ask the boy on what day of the week he was born.

Assume that which of the two children answers the door is determined by chance. Then the procedure was (1) pick a two-child family at random from all two-child families (2) pick one of the two children at random, (3) see if it is a boy and ask on what day he was born. The chance the other child is a girl is $\frac{1}{2}$. This is a very different procedure from (1) picking a two-child family at random from all families with two children, at least one a boy, born on a Tuesday. The chance the family consists of a boy and a girl is $\frac{14}{27}$, about 0.52.

This variant of the boy and girl problem is discussed on many internet blogs and is the subject of a paper by Ruma Falk.^[14] The moral of the story is that these probabilities do not just depend on the known information, but on how that information was obtained.

Psychological investigation

From the position of statistical analysis the relevant question is often ambiguous and as such there is no "correct" answer. However, this does not exhaust the boy or girl paradox for it is not necessarily the ambiguity that explains how the intuitive probability is derived. A survey such as vos Savant's suggests that the majority of people adopt an understanding of Gardner's problem that if they were consistent would lead them to the $\frac{1}{3}$ probability answer but overwhelmingly people intuitively arrive at the $\frac{1}{2}$ probability answer. Ambiguity notwithstanding, this makes the problem of interest to psychological researchers who seek to understand how humans estimate probability.

Fox & Levav (2004) used the problem (called the *Mr. Smith problem*, credited to Gardner, but not worded exactly the same as Gardner's version) to test theories of how people estimate conditional probabilities.^[2] In this study, the paradox was posed to participants in two ways:

- "Mr. Smith says: 'I have two children and at least one of them is a boy.' Given this information, what is the probability that the other child is a boy?"
- "Mr. Smith says: 'I have two children and it is not the case that they are both girls.' Given this information, what is the probability that both children are boys?"

The authors argue that the first formulation gives the reader the mistaken impression that there are two possible outcomes for the "other child",^[2] whereas the second formulation gives the reader the impression that there are four possible outcomes, of which one has been rejected (resulting in $\frac{1}{3}$ being the probability of both children being boys, as there are 3 remaining possible outcomes, only one of which is that both of the children are boys). The study found that 85% of participants answered $\frac{1}{2}$ for the first formulation, while only 39% responded that

way to the second formulation. The authors argued that the reason people respond differently to each question (along with other similar problems, such as the [Monty Hall Problem](#) and the [Bertrand's box paradox](#)) is because of the use of naive [heuristics](#) that fail to properly define the number of possible outcomes.^[2]

See also

- [Monty Hall problem](#)
- [Necktie paradox](#)
- [Sleeping Beauty problem](#)
- [St. Petersburg paradox](#)
- [Two envelopes problem](#)

References

1. Martin Gardner (1954). *The Second Scientific American Book of Mathematical Puzzles and Diversions*. Simon & Schuster. ISBN 978-0-226-28253-4.
2. Craig R. Fox & Jonathan Levav (2004). "Partition–Edit–Count: Naive Extensional Reasoning in Judgment of Conditional Probability". *Journal of Experimental Psychology*. **133** (4): 626–642. doi:10.1037/0096-3445.133.4.626. PMID 15584810.
3. Martin Gardner (1961). *The Second Scientific American Book of Mathematical Puzzles and Diversions*. Simon & Schuster. ISBN 978-0-226-28253-4.
4. Maya Bar-Hillel and Ruma Falk (1982). "Some teasers concerning conditional probabilities". *Cognition*. **11** (2): 109–122. doi:10.1016/0010-0277(82)90021-X. PMID 7198956.
5. [Raymond S. Nickerson](#) (May 2004). *Cognition and Chance: The Psychology of Probabilistic Reasoning*. Psychology Press. ISBN 0-8058-4899-1.
6. "Ask Marilyn". Parade Magazine. October 13, 1991 [January 5, 1992; May 26, 1996; December 1, 1996; March 30, 1997; July 27, 1997; October 19, 1997].
7. Tierney, John (2008-04-10). "The psychology of getting suckered". The New York Times. Retrieved 24 February 2009.
8. Leonard Mlodinow (2008). *The Drunkard's Walk: How Randomness Rules our Lives*. Pantheon. ISBN 0-375-42404-0.
9. Nikunj C. Oza (1993). "On The Confusion in Some Popular Probability Problems". [CiteSeerX](#) 10.1.1.44.2448.
10. P.J. Laird; et al. (1999). "Naive Probability: A Mental Model Theory of Extensional Reasoning". *Psychological Review*. **106**: 62–88. doi:10.1037/0033-295x.106.1.62.
11. Matthew A. Carlton and William D. Stansfield (2005). "Making Babies by the Flip of a Coin?". *The American Statistician*. **59**: 180–182. doi:10.1198/000313005x42813.
12. Charles M. Grinstead and J. Laurie Snell. "Grinstead and Snell's Introduction to Probability" (PDF). The CHANCE Project.
13. Stephen Marks and Gary Smith (Winter 2011). "The Two-Child Paradox Reborn?" (PDF). *Chance (Magazine of the American Statistical Association)*. **24**: 54–9. doi:10.1007/s00144-011-0010-0.
14. [1] (<http://www.tandfonline.com/doi/abs/10.1080/13546783.2011.613690>)

External links

- Boy or Girl: Two Interpretations (<http://mathforum.org/library/drmath/view/52186.html>)
- At Least One Girl (<http://www.mathpages.com/home/kmath036/kmath036.htm>) at MathPages
- A Problem With Two Bear Cubs (<http://www.cut-the-knot.org/bears.shtml>)
- Lewis Carroll's Pillow Problem (<http://www.cut-the-knot.org/carroll.shtml>)
- When intuition and math probably look wrong (http://www.sciencenews.org/view/generic/id/60598/title/Math_Trek_When_intuition_and_math_probably_look_wrong)
- The Boy or Girl Paradox: A Martingale Perspective (<http://finmathbloG·Blogspot.com/2013/07/the-boy-or-girl-paradox-martingale.html>) at FinancialMathematics.com

Retrieved from "https://en.wikipedia.org/w/index.php?title=Boy_or_Girl_paradox&oldid=766674814"

Categories: [Probability theory paradoxes](#) | [Probability problems](#)

This page was last modified on 21 February 2017, at 14:11.

Text is available under the [Creative Commons Attribution-ShareAlike License](#); additional terms may apply. By using this site, you agree to the [Terms of Use](#) and [Privacy Policy](#). Wikipedia® is a registered trademark of the [Wikimedia Foundation, Inc.](#), a non-profit organization.

- [Contact Wikipedia](#)
- [Developers](#)
- [Cookie statement](#)