Revisiting the Name Variant of the Two-Children Problem

Davy Paindaveine\textsuperscript{a} and Philippe Spindel\textsuperscript{b,c}

\textsuperscript{a}ECARES and Department of Mathematics, Université libre de Bruxelles, Bruxelles, Belgium; \textsuperscript{b}Service de Physique de l’Univers, Champs et Gravitation, Université de Mons, Mons, Belgium; \textsuperscript{c}Service de Physique Théorique, Université libre de Bruxelles, Bruxelles, Belgium

\textbf{ABSTRACT}

Initially proposed by Martin Gardner in the 1950s, the famous two-children problem is often presented as a paradox in probability theory. A relatively recent variant of this paradox states that, while in a two-children family for which at least one child is a girl, the probability that the other child is a boy is 2/3, this probability becomes 1/2 if the first name of the girl is disclosed (provided that two sisters may not be given the same first name). We revisit this variant of the problem and show that, if one adopts a natural model for the way first names are given to girls, then the probability that the other child is a boy may take any value in (0, 2/3). By exploiting the concept of Schur-concavity, we study how this probability depends on model parameters.

\textbf{1. Introduction}

In one of his famous mathematical games, Martin Gardner asked the following questions: Mr. Smith has two children. At least one of them is a boy. What is the probability that the other child is a girl? Mr. Jones has two children. The older is a girl. What is the probability that the other child is a girl? See Gardner (1959). While the answers he first provided were 1/3 and 1/2, respectively, he explicitly stated later that the first question was actually ambiguous; see Gardner (1987), chap. 14 and 19. To phrase this question without ambiguity in such a way that the corresponding answer is indeed 1/3, one should adopt the view (as we do throughout the present note) that Mr. Smith is randomly selected among all two-children families having at least one boy; see Bar-Hillel and Falk (1982) or Khovanova (2012) for discussions on how the answer to this first question depends on the way one obtains the information that Mr. Smith has at least one boy.

The two-children problem, that is sometimes referred to as the boy-or-girl problem, gained much popularity twenty-five years ago when it was discussed by the well-known columnist Marilyn vos Savant in \textit{Parade} magazine (vos Savant 1996). It has since been discussed in several monographs (Mlodinow 2008; Chang 2012) and scientific papers (among which D’Agostini 2010; Lynch 2011; Pollak 2013, and the aforementioned ones).

Two variants of the two-children problem are famous. The first one asks: for a two-children family having at least a girl who is born on a Tuesday, what is the probability that the other child is a boy? See Lynch (2011), Falk (2011), Taylor and Stacey (2014), and Zaskis and Wijeratne (2015). If the probability that a girl is born on a Tuesday is \( r \), then this probability can be shown to be equal to \( 2/(4 - r) \), which ranges from 1/2 (for \( r = 0 \)) to 2/3 (for \( r = 1 \)); for \( r = 1/7 \), the probability that the other child is a boy is thus 14/27. This still assumes that the two-children family considered here is randomly selected from all two-children families meeting this particular property; when removing the ambiguity above in another way, different probabilities are obtained in this variant, too; see Khovanova (2012).

The second variant rather asks: for a two-children family having at least a girl whose name\textsuperscript{1} is Florida, what is the probability that the other child is a boy? See, for example, Mlodinow (2008) or Marks and Smith (2011). If two sisters may be given the same name, then this variant is strictly equivalent to the previous one: more precisely, if it is assumed that girls are independently named Florida with probability \( r \), then the probability that the other child is a boy is \( 2/(4 - r) \). To make the second variant of interest, one therefore needs to assume that two sisters may not be given the same name, in which case, under the assumptions associated with what we will call Model A below, the probability that the other child is a boy is \( 1/2 \), irrespective of \( r \); see D’Agostini (2010).

In this note, we revisit this second variant. In Section 2, we describe two models that can be considered to answer the question of interest. The first one, Model A, is the traditional one, for which both genders are equally likely for the other child—we carefully state the corresponding assumptions. Then, we introduce a new model, Model B, that specifies the way girl names are picked by parents. In Section 3, we compute in each model the probability that the other child is a boy. While both genders are indeed equally likely in Model A, this probability in Model B may assume any value in (0, 2/3) depending on popularities of girl names. In this second model, this probability is actually a Schur-concave function of name popularities, which allows us to study how this probability depends on model parameters. In Section 4, we provide some final comments. Finally, an Appendix collects technical proofs.

\textsuperscript{1}In the rest of the article, “name” will throughout stand for “first name”.

\textsuperscript{a}ECARES and Department of Mathematics, Université libre de Bruxelles, Bruxelles, Belgium.

© 2023 American Statistical Association
2. Two Models

We consider two probabilistic models, labeled Model A and Model B below, for the gender and name of each child in a two-children family (in line with the question raised in the second variant of the two-children problem, we will actually consider names for girls only). We start by describing assumptions that are common to both models. Regarding gender, we adopt the following assumption.

Assumption 1 (Models A–B). (i) Any born child is a boy (b) with probability 1/2 and a girl (g) with probability 1/2. (ii) Genders of both children are independent.

Writing $E$ and $Y$ for Elder and Younger, respectively, we thus have, with obvious notation, $P[E|b] = P[E|g] = P[Y|b] = P[Y|g] = 1/2$, hence, for example, $P[E|b \cap Y|b] = 1/4$. Now, we turn to assumptions involving names. We let $r_1 := P[E_1|E_2]$, where $E_2$ is the event that the elder child is a girl named $n_1$. This event has thus probability $P[E_2] = P[E_2|E_1]P[E_1] = r_1/2$, which yields $P[E \setminus E_1] = 1 - P[E|b] - P[E|g] = (1 - r_1)/2$. The assumptions common to both models and related to names are then as follows.

Assumption 2 (Models A–B). (i) Two sisters may not be given the same name: $P[E_1 \cap Y_1] = 0$. (ii) The name given to an elder girl and gender of the second child are independent: $P[Y_1|E_1] = P[Y_1|E_2] = (1/2)$. (iii) A girl having an elder brother is given name $n_1$ with the same probability as an elder girl child: $P[Y_1|(Y \cap E_2)] = P[E_2|E_1](= r_1)$.

Assumption 2(ii) yields $P[E_2 \cap Y_2] = P[Y_2|E_1]P[E_2] = r_1/4$, which implies both $P[\bar{E} \setminus E_2 \cap Y_2] = Y_2 - P[E_2 \cap Y_2] - P[E_2|Y_2] = r_1/4$ and (using also Assumption 2(i))

$P[E_2 \cap (Y_2 \setminus E_2)] = P[E_2] - P[E_2 \setminus Y_2] - (1 - r_1)/4$. Moreover, Assumption 2(iii) provides $P[E_2 \setminus Y_2] = P[Y_2|E_2 \setminus Y_2] = (1 - r_1)/4$. Summing up, Assumptions 1–2 lead to the probabilities given in Table 1. Clearly, one needs an extra assumption to determine the missing probabilities in this table, and the two models will differ.

Assumption 3 (Model A). $P[Y_1|E_1] = P[E_1|E_1](= r_1)$, or equivalently $P[Y_1] = P[E_1](= r_1/2)$.

From a statistical point of view, this modeling assumption essentially translates the expectation that, within two-children families, there should be roughly as many girls baring the name $n_1$ among the younger女孩 children as among elder girl ones. Under this assumption, we indeed have $p_{32} = P[Y_1|E_1] = P[E_1](= r_1/2)$, which allows us to obtain $p_{32} = r_1/4$, $p_{33} = (1 - 2r_1)/4$ and $p_3 = (1 - r_1)/2$. Note that this imposes that $r_1 \leq 1/2$, a restriction we will not have in the alternative model we now describe.

Unlike the model above, Model B does not rely on a statistical view but rather provides a probabilistic scheme describing how girl names are picked by parents according to popularity. Assume that there are $K$ names, $n_1, \ldots, n_K$, say, with respective popularity $r_1, \ldots, r_K$, where the $r_k$’s are positive numbers that sum up to one. If the elder child is a girl, then it will be accordingly given name $n_k$ with probability $r_k$. If the first child is a boy and the second one is a girl, then the $K$ names are available for this girl, which will similarly be given name $n_k$ with probability $r_k$. However, if the first child is a girl, named $n_k$ say, then in case a second girl is born, this name is not available anymore (Assumption 2(i)), and parents will then naturally give this girl a name according to the rescaled probabilities associated with $r_1, \ldots, r_{K-1}, r_{K+1}, \ldots, r_K$. This is formalized in the following assumption.

Assumption 3 (Model B). Girl names available are $n_1, \ldots, n_K$. The first girl born in a family (if any) is given name $n_k$ with probability $r_k$; here, $r_k \geq 0$ for any $k = 1, \ldots, K$ and $\sum_{k=1}^{K} r_k = 1$. If the elder girl was given name $n_k$, then the possible second girl is given name $n_\ell$, with probability $r_\ell (1 - \delta_{k\ell})/(1 - r_k)$, where $\delta_{k\ell}$ takes value one if $k = \ell$ and value zero otherwise: $P[Y_g|(E_2, Y_2, K)] = r_\ell (1 - \delta_{k\ell})/(1 - r_k)$. Consequently, Model B yields $p_{32} = \sum_{k=1}^{K} P[E_2|Y_2, K]$, hence,

$$p_{32} = \frac{r_1}{4} + \sum_{k=2}^{K} \frac{r_k}{4} \left(1 + \sum_{k=2}^{K} \frac{r_k}{1 - r_k}\right).$$

Of course, one then has $p_{33} = (1 - r_1)/4 - p_{32}$ and $p_3 = (1/2) - p_{32}$, but these values will not be needed for our purposes. Note that, in contrast with Model A, here $p_3 = P[Y_1](= r_1/2)$. We argue that this is natural in the setup considered: while names are obviously picked by parents, it seems spurious to assume that the mechanism they adopt to choose names will ensure that $P[Y_1] = P[E_1]$. For instance, it turns out that $n_1$ is a very popular name, then one would expect that $P[Y_1] < P[E_1]$, since parents in two-daughters families are likely to pick this popular name for the first child. Another advantage of Model B is that arbitrarily high popularity is allowed: $r_1$ may assume any value in (0, 1) in this model, whereas the constraint $P[Y_1] = P[E_1]$ inherent to Model A excludes that $r_1$ exceeds 1/2. Allowing for arbitrarily high popularity of a name is natural and attractive from a mathematical point of view (if not from a practical point of view).

Table 1. Probabilities obtained from Assumptions 1–2.

|      | $Y_b$ | $Y_g$ | $Y_b \setminus Y_g$ |
|------|------|------|---------------------|
| $E_b$ | 1/4  |      |                     |
| $E_g$ |      | $r_1/4$ | $(1 - r_1)/4$       |
| $E_g \setminus E_1$ | $(1 - r_1)/4$ | $p_{32}$ | $p_{33}$ | $(1 - r_1)/2$ |

$p_{32} = \frac{r_1}{4}$ and $p_{33} = (1 - 2r_1)/4$ and $p_3 = (1 - r_1)/2$. Note that this imposes that $r_1 \leq 1/2$, a restriction we will not have in the alternative model we now describe.
3. Results

In any model satisfying Assumptions 1–2, the probability that a family has a boy given that it has a girl named \( n_1 \) is

\[
P(\text{Eb} \cup \text{Yb}|\text{Eg}_1 \cup \text{Yg}_1) = \frac{P(\text{Eb} \cup \text{Yb} \cap (\text{Eg}_1 \cup \text{Yg}_1))}{P(\text{Eg}_1 \cup \text{Yg}_1)} = \frac{(r_1/4) + (r_1/4)}{(r_1/2) + p_2^K};
\]

see Table 1. In Model A, we have \( p_2 = p_2^A = r_1/2 \), which yields

**Theorem 1 (Model A).** The probability that a family has a boy given that it has a girl named \( n_1 \) is \( P(\text{Eb} \cup \text{Yb}|\text{Eg}_1 \cup \text{Yg}_1) = 1/2 \), irrespective of the value of \( r_1 \).

This is the usual result that, if a family has a girl named \( n_1 \) (and if two sisters may not be given the same name), then both genders are equally likely for the other child. We now turn to Model B, for which the situation is very much different. Up to renumbering the names \( n_2, \ldots, n_K \), we may of course assume that \( r_2 \geq r_3 \geq \cdots \geq r_K \) (note that this does not impose anything on \( r_1 \)). The model is thus indexed by \( R_K = \{ r = (r_1, r_2, \ldots, r_K) \in (0, 1) : \sum^K_{k=1} r_k = 1, r_2 \geq r_3 \geq \cdots \geq r_K \} \). For any \( r \in R_K \), the probability that a family has a boy given that it has a girl named \( n_1 \) is then

\[
P(\text{Eb} \cup \text{Yb}|\text{Eg}_1 \cup \text{Yg}_1) = \frac{(r_1/4) + (r_1/4)}{(r_1/2) + p_2^B} = \frac{2}{3 + \sum^K_{k=2} \frac{r_k}{1 - r_k}}.
\]

As \( r \to (0, 1, 0, \ldots, 0) \), this probability converges to zero. Since \( r_k > 0 \) for any \( k \), it trivially holds that this probability is strictly smaller than 2/3, an upper bound that is obtained as \( r \to (1, 0, \ldots, 0) \). From continuity, the probability in (1) can take any value in the interval \((0, \frac{2}{3})\). We therefore proved the following result.

**Theorem 2 (Model B).** Fix an integer \( K \geq 2 \). Then, for any \( r \in R_K \), the probability that the family has a boy given that it has a girl named \( n_1 \) is

\[
p(r) = \frac{2}{3 + \sum^K_{k=2} \frac{r_k}{1 - r_k}}.
\]

Moreover, \( p(R_K) \), the image of \( R_K \) under the mapping \( p \), is \((0, \frac{2}{3})\).

For any \( K \geq 2 \), the uniform configuration \( r = (1/K, \ldots, 1/K) \) provides the case \( p(r) = 1/2 \) where both genders are equally likely for the other child. Figure 1 plots \( p(r) \) as a function of \( r_1 \) for various values of \( K \) in the framework where the names \( n_2, \ldots, n_K \) are equally likely (the figure also provides Monte Carlo simulation results that clearly support our expression of \( p(r) \) in Theorem 2). It will be of interest to determine for which values of \( r_1 \) there exists a corresponding configuration \( r = (r_1, \ldots, r_K) \) such that the probability \( p(r) \) that the other child is a boy, that is, for such \( K \) and \( r_1 \), this probability will much depend on the popularities of the remaining \( K - 1 \) names. Note that the upper bound in (2) reduces to \( 2r_1/(2r_1 + 1) \) for \( K = 2 \). As expected, for any \( r_1 \in (0, 1) \), the (closures of the) feasible sets \( p(R_K(r_1)) \), \( K = 2, 3, \ldots \), form a strictly increasing sequence with respect to inclusion, and

\[
\lim_{K \to \infty} p(R_K(r_1)) = \left( \frac{2r_1}{2r_1 + 1} \right) \left( \frac{2}{4 - r_1} \right).
\]

**Theorem 3 (Model B).** Fix an integer \( K \geq 2 \) and define \( S_K := \{ r_1 \in (0, 1) : p(R_K(r_1)) \geq \frac{1}{2} \} \), the collection of \( r_1 \)-values for which some configuration \( r = (r_1, \ldots, r_K) \) makes, in a family

![Figure 1](https://via.placeholder.com/150)
having a girl named $n_1$, both genders equally likely for the other child. Then,

$$
S_2 = \begin{bmatrix} 1 \\
2
\end{bmatrix}
$$

and

$$
S_K = \begin{bmatrix} \frac{1}{K} \\
2
\end{bmatrix}
$$

for $K > 2$.

As mentioned above, for any $K \geq 2$, the uniform case $r_1 = r_2 = \cdots = r_K (= 1/K)$ provides $p(r) = 1/2$. For $K = 2$, this is the only case leading to equally likely genders for the second child. For $K = 3$, it is readily checked that the cases $r = (r_1, r_2, r_3) \in \mathcal{R}_3$ providing $p(r) = 1/2$ are described by

$$
r_2 = \frac{1 - r_1}{2} + \frac{(3r_1 + 1)^2 - 4}{6} \quad \text{and} \quad
r_3 = \frac{1 - r_1}{2} - \frac{(3r_1 + 1)^2 - 4}{6},
$$

with $r_1 \in [1/3, 1/2]$. Figure 3 offers a graphical representation. Note that the results are in line with Corollary 3.1. More generally, for $K \geq 4$, the collections of $(r_2, \ldots, r_K)$ making both genders equally likely for the second child is a manifold of dimension $K - 3$ in $\mathbb{R}^{K-1}$.

### 4. Final Comments

This article revisits the name variant of the two-children problem and shows that, if it is known that the family has a girl named $Florida$, say, then the probability that the other child is a boy may depend on the stochastic model that is adopted, even if one restricts to models that do not allow two sisters to have the same name. We show this by recalling that this probability is always $1/2$ in the traditional model and by defining an alternative, natural, model in which this probability may assume any value in

$$
\left[ \frac{2r_1}{2r_1 + 1}, \frac{2(K - 2 + r_1)}{4(K - 2 + r_1) + 1 - Kr_1} \right],
$$

where $r_1$ is the popularity of the name $Florida$ (more precisely, $r_1$ is the probability that the first girl born in a family is given this name) and $K$ is the number of possible names (this assumes that $K \geq 3$; for $K = 2$, this probability must be equal to $2r_1/(2r_1 + 1)$). In this new model, the dependence of this probability on $r_1$ clearly results from the heterogeneous way names are assigned to a first girl or a second one in a family.

It is of course natural to investigate whether or not the assumptions from Section 2 are valid. For Assumption 1, this was discussed in Carlton and Stansfield (2005), where the authors conclude that neither Part (i) nor Part (ii) of this assumption actually holds in practice: more precisely, births of boys are more frequent than births of girls, and independence of genders is violated; in line with this, the null hypothesis that the number of boys follows a binomial distribution with parameters 2 and $p$ for some unspecified $p \in [0, 1]$ is rejected at all usual significance levels (the $p$-value is below $10^{-5}$). While results of the present note can be trivially extended to asymmetric gender probabilities, it is unclear how to deal with violation of the independence assumption (to the best of the authors’ knowledge, this issue actually has not been touched for the classical Model A). While it would also be of interest to test whether or not Assumption 2 holds in practice, it would be more urgent, in the context of the present article, to focus on Assumption 3 and to investigate which model, among Model A and Model B, provides a better description of the real world. This could be done, on the basis of suitable data, by studying whether or not the proportion of girls named $n_1$ among elder girls is different from the corresponding proportion among
Appendix: Technical Proofs

In this appendix, we prove Theorem 3 and Corollary 3.1.

Proof of Theorem 3. The result for \( K = 2 \) trivially follows from Theorem 2. Fix then \( K > 2 \) and \( r_1 \in (0, 1) \). Since the function \( x \mapsto g(x) = x/(1 - x) \) is strictly convex over \((0, 1)\), Proposition C.1a in p. 92 from Marshall et al. (2011) yields that the function

\[
(r_2, \ldots, r_K) \mapsto \frac{\sum_{k=2}^{K} r_k}{1 - r_k}
\]

is strictly Schur-convex on \((0, 1)^{K-1}\), so that

\[
(r_2, \ldots, r_K) \mapsto p(r_1, r_2, \ldots, r_K) = \frac{2}{3 + \sum_{k=2}^{K} \frac{r_k}{1 - r_k}}
\]

is strictly Schur-concave on \((0, 1)^{K-1}\). For any \( r \in \mathcal{R}_K(r_1) \), we have \( \sum_{k=2}^{K} r_k = 1 - r_1 \), hence, also

\[
\left( \frac{1 - r_1}{K - 1}, \ldots, \frac{1 - r_1}{K - 1} \right) < (r_2, \ldots, r_K) < (1 - r_1, 0, \ldots, 0),
\]

where \( \prec \) refers to the usual majorization ordering; see, for example, pp. 8–9 in Marshall et al. (2011). The above Schur-concavity therefore implies that

\[
p(r_1, 1 - r_1, 0, \ldots, 0) \leq p(r) \leq p\left( r_1, \frac{1 - r_1}{K - 1}, \ldots, \frac{1 - r_1}{K - 1} \right)
\]

for any \( r \in \mathcal{R}_K(r_1) \). Since direct evaluation provides

\[
p(r_1, 1 - r_1, 0, \ldots, 0) = \frac{2r_1}{2r_1 + 1}
\]

and

\[
p\left( r_1, \frac{1 - r_1}{K - 1}, \ldots, \frac{1 - r_1}{K - 1} \right) = \frac{2(K - 2 + r_1)}{4(K - 2 + r_1) + 1 - Kr_1},
\]

the result follows from continuity of \( r \mapsto p(r) \). □

Proof of Corollary 3.1. The result trivially holds for \( K = 2 \), so we may restrict to \( K > 2 \). From Theorem 3, we have that \( r_1 \in S_K \)—that is, \( \frac{1}{2} \notin p(\mathcal{R}_K(r_1)) \)—if and only if

\[
\frac{2r_1}{2r_1 + 1} < \frac{1}{2} \quad \text{and} \quad \frac{1}{2} \leq \frac{2(K - 2 + r_1)}{4(K - 2 + r_1) + 1 - Kr_1},
\]

that is, if and only if

\[
r_1 < \frac{1}{2} \quad \text{and} \quad r_1 \geq \frac{1}{K},
\]

which establishes the result. □

Acknowledgments

The authors would like to thank the Editor, the Associate Editor and two referees for the careful reviews of the manuscript and insightful comments and suggestions. The present work results from exchanges following a talk of the Altair conference cycle in Brussels; we would like to thank the organizers.

Disclosure Statement

The authors report there are no competing interests to declare.

Funding

This research is supported by the Program of Concerted Research Actions (ARC) of the Université libre de Bruxelles and by the Fonds Thelam from the Fondation Roi Baudouin.

References

Bar-Hillel, M., and Falk, R. (1982), "Some Teasers Concerning Conditional Probabilities," Cognition, 11, 109–122. [401]

Carlton, M., and Stansfield, W. (2005), "Making Babies by the Flip of a Coin?" The American Statistician, 59, 180–182. [404,405]

Chang, M. (2012), Paradoxes in Scientific Inference, New York: Chapman and Hall/CRC. [401]

D’Agostini, G. (2010), "On the So Called Boy or Girl Paradox," arXiv:1001.0708v1. [401]

Falk, R. (2011), "When Truisms Clash: Coping with a Counterintuitive Problem Concerning the Notorious Two-Child Family," Thinking & Reasoning, 17, 353–366. [401]

Gardner, M. (1959), Mathematical Games," Scientific American, 200, 164–174. [401]

--- (1987), "The Second Scientific American Book of Mathematical Puzzles and Diversions, Chicago: The Chicago University Press. [401]

Khovanova, T. (2012), "Martin Gardner’s Mistake," The College Mathematics Journal, 43, 20–24. [401]

Lynch, P. (2011), "The Two-Child Paradox: Dichotomy and Ambiguity," Irish Mathematical Society Bulletin, 67, 67–73. [401]

Marks, S., and Smith, G. (2011), "The Two-Child Paradox Reborn?" Chance, 24, 54–59. [401]

Marshall, A. W., Olkin, I., and Arnold, B. C. (2011), Inequalities: Theory of Majorization and Its Applications (2nd ed.), New York: Springer. [405]

Mlodinow, L. (2008), The Drunkard’s Walk: How Randomness Rules Our Lives, New York: Pantheon Books. [401]

Pollak, M. (2013), "A Stochastic Dominance Property Common to the Boy-or-Girl Paradox and the Lottery," Statistics and Probability Letters, 83, 410–413. [401]

Taylor, W., and Stacey, K. (2014), "Gardner’s Two Children Problems and Variations: Puzzles with Conditional Probability and Sample Spaces," The Australian Mathematics Teacher, 70, 1–19. [401]

vos Savant, M. (1996), "Ask Marilyn," Parade Magazine, May 26, 17. [401]

Zaksis, R., and Wijeratne, C. (2015), "Two Boys Problem Revisited, or, 'What has Tuesday got to do with it?”, Mathematics Teaching, 18–21. [401]