



CORRESPONDENCE

**REVISED** Identical twins and Bayes' theorem in the 21st century  
[version 2; peer review: 2 not approved]

Valentin Amrhein<sup>1,2</sup>, Tobias Roth<sup>1,2</sup>, Fränzi Korner-Nievergelt<sup>3,4</sup>

<sup>1</sup>Zoological Institute, University of Basel, 4051 Basel, Switzerland  
<sup>2</sup>Research Station Petite Camargue Alsacienne, 68300 Saint-Louis, France  
<sup>3</sup>Oikostat GmbH, 6218 Ettiswil, Switzerland  
<sup>4</sup>Swiss Ornithological Institute, 6204 Sempach, Switzerland

**v2** **First published:** 17 Dec 2013, 2:278 (<https://doi.org/10.12688/f1000research.2-278.v1>)  
**Latest published:** 29 Jul 2015, 2:278 (<https://doi.org/10.12688/f1000research.2-278.v2>)

**Abstract**  
 In an article in *Science* on "Bayes' Theorem in the 21st Century", Bradley Efron uses Bayes' theorem to calculate the probability that twins are identical given that the sonogram shows twin boys. He concludes that Bayesian calculations cannot be uncritically accepted when using uninformative priors. While we agree that the choice of the prior is essential, we argue that the calculations on identical twins give a biased impression of the influence of uninformative priors in Bayesian data analyses.

**Keywords**  
 Baye's theorem, identical twins, Bayesian, uninformative priors

**Open Peer Review**

**Reviewer Status** XX

	Invited Reviewers	
	1	2
<b>REVISED</b>	<span style="color: red;">X</span>	<span style="color: red;">X</span>
<b>version 2</b> published 29 Jul 2015	report	report
	↑	
<b>version 1</b> published 17 Dec 2013	<span style="color: red;">X</span>	
	report	

---

1 **Michael McCarthy**, University of Melbourne, Melbourne, Australia

2 **Andrew Gelman**, Columbia University, New York, NY, USA

Any reports and responses or comments on the article can be found at the end of the article.

**Corresponding author:** Valentin Amrhein ([v.amrhein@unibas.ch](mailto:v.amrhein@unibas.ch))

**Competing interests:** No competing interests were disclosed.

**Grant information:** This work was funded by the Swiss Association Pro Petite Camargue Alsacienne and the Fondation de Bienfaisance Jeanne Lovioz.  
*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Copyright:** © 2015 Amrhein V *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

**How to cite this article:** Amrhein V, Roth T and Korner-Nievergelt F. **Identical twins and Bayes' theorem in the 21st century [version 2; peer review: 2 not approved]** F1000Research 2015, 2:278 (<https://doi.org/10.12688/f1000research.2-278.v2>)

**First published:** 17 Dec 2013, 2:278 (<https://doi.org/10.12688/f1000research.2-278.v1>)

**REVISED** Amendments from Version 1

In our manuscript, we now clarified that our approach is different from the calculations provided by Efron. We also shortened the manuscript and removed statements that were criticized by referee Michael McCarthy.

**See referee reports**

## Correspondence

Efron<sup>1</sup> provides four examples of Bayesian analyses, two of which underline the remarkable potential of Bayesian methods. Based on one of the other examples, however, Efron ultimately concludes that Bayesian analyses using uninformative priors cannot be uncritically accepted and should be checked by frequentist methods. While we wholeheartedly agree that statistical results should not be uncritically accepted, we find Efron's example ineffective in showing that Bayesian statistics require more careful checking than any other kind of statistics.

In his example on uninformative priors, Efron uses Bayes' theorem to calculate the probability that twins are identical given that the sonogram shows twin boys. Efron finds this probability to be 2/3 when using an uninformative prior versus 1/2 with an informative prior and thereby concludes that an uninformative prior does not have the desired neutral effects on the output of Bayes' rule. We argue that this example is relatively useless in illustrating Bayesian data analysis. One reason is that Efron considers the particular set of twin boys as the entire population. In this case, statistics is not needed because there is no random sample drawn from a larger population. Rather, Efron combines different pieces of expert knowledge from the doctor and genetics using Bayes' theorem. While certainly an impeccable probability law, Bayes' theorem is a mathematical equation, not a statistical model describing how data may be produced. In essence, Efron uses this equation to show that the value on the left side of the equation changes when a term on the right side is changed, which is trivial and could be shown with any mathematical equation also in a non-Bayesian context.

Efron's example can be rearranged so that it fits a more realistic situation in statistical data analysis, albeit with a very low sample size: consider the twin boys that, as Efron casually mentions, turned out to be fraternal, as a random sample from the larger population of twin boys and try to draw inference about the proportion of identical twins among the population of twin boys (note that this approach is different from the calculations provided by Efron). If we use the data point together with an uninformative uniform prior on  $P(A|B)$  (see [Box 1](#)) to determine the probability of identical twins given the twins are two boys, we obtain, with 95% certainty, a probability of between 0.01 and 0.84; if we use a highly informative prior based on information from the doctor and genetics, we obtain a probability of between 0.49 and 0.51. This looks completely reasonable to us, although of course we do not know much more than

### Box 1. Study question: What is the probability of identical twins given the twins are two boys?

Data: One pair of twin boys is fraternal.

Data model:  $x \sim \text{Binomial}(\theta, n)$ , where  $\theta$  is the probability of identical twins given the twins are two boys,  $x$  is the number of identical twins in the data, and  $n$  is the total number of pairs of twin boys; in our case:  $x = 0$  and  $n = 1$ .

The posterior distribution  $p(\theta|x)$  is obtained using Bayes' theorem

$$p(\theta|x) = p(x|\theta)p(\theta)/p(x)$$

We use two different priors  $p(\theta)$ :

1) Uninformative prior:  $p(\theta) = \text{Unif}(0,1) = \text{Beta}(1,1)$

2) Informative prior: using the information from the doctor and from genetics, we are quite sure that  $\theta$  must be around 0.5<sup>1</sup> Transforming this information into a statistical distribution yields  $p(\theta) = \text{Beta}(10000, 10000)$ , which has a mean of 0.5 and a 95% interval of 0.493 – 0.507. [Note that we had to choose the 95% interval arbitrarily because we are not informed about the certainty of the information provided by the doctor and by genetics].

Given the single parameter Binomial model,  $x \sim \text{Binomial}(\theta, n)$ , and the prior  $p(\theta) = \text{Beta}(\alpha, \beta)$ , the solution of the Bayesian analysis is given by the posterior distribution  $p(\theta|x) = \text{Beta}(\alpha+x, \beta+n-x)$  [see any Bayesian textbook, e.g. Gelman *et al.* 2004<sup>2</sup>, p. 34]

The probability of identical twins given the twins are two boys:

1) Uninformative prior:  $p(\theta|x) = \text{Beta}(1+x, 1+n-x) = \text{Beta}(1+0, 1+1-0) = \text{Beta}(1, 2)$ , which has an expected value of 0.33 and a 95% interval of 0.013 – 0.84.

2) Informative prior:  $p(\theta|x) = \text{Beta}(10000+x, 10000+n-x) = \text{Beta}(10000+0, 10000+1-0) = \text{Beta}(10000, 10001)$ , which has an expected value of 0.50 and a 95% interval of 0.49 – 0.51.

we knew before because we had only a single data point. We think that to illustrate the influence of non-informative priors on results of Bayesian data analyses, such an approach would be fairer than the calculations given by Efron.

Although we agree with Efron<sup>1</sup> that the choice of the prior is essential, we conclude that his article gives a biased impression of the influence of uninformative priors. In his example using Bayes' theorem, we found no reliable support for his main conclusion that Bayesian calculations cannot be uncritically accepted when using uninformative priors.

### Author contributions

FK-N analyzed the data point. VA wrote the first draft of the manuscript. All authors contributed to the discussion and approved the final version of the manuscript.

### Competing interests

No competing interests were disclosed.

### Grant information

This work was funded by the Swiss Association Pro Petite Camargue Alsacienne and the Fondation de Bienfaisance Jeanne Lovioz.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

### Acknowledgements

The authors would like to thank Yves-Laurent Grize and Pius Korner for discussions, and Michael McCarthy for valuable comments on the manuscript.

## References

---

1. Efron B: **Bayes' Theorem in the 21st Century**. *Science*. 2013; **340**(6137): 1177–1178.  
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Gelman A, Carlin JB, Stern HS, *et al.*: **Bayesian Data Analysis**. Chapman & Hall, New York. 2004.  
[Publisher Full Text](#)

# Open Peer Review

Current Peer Review Status: **×** **×**

---

## Version 2

Reviewer Report 07 November 2017

<https://doi.org/10.5256/f1000research.7373.r11175>

© 2017 Gelman A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



### Andrew Gelman

Department of Statistics, Columbia University, New York, NY, USA

I don't think the analysis in this paper, or that of Efron, is correct. We actually compute  $\Pr(\text{identical twins | twin brother})$  as an exercise in chapter 1 of Bayesian Data Analysis (originally published in 1995); we estimate the probability as 5/11 (see solution to exercise 1.6 here:

<http://www.stat.columbia.edu/~gelman/book/solutions.pdf>). The probability has surely gone down in recent years with the rise in popularity of fertility treatments, which have drastically increased the rate of non-identical twins. So I don't think that either 2/3 or 1/2 is a good answer!

Maybe I'm missing something here?

**Is the rationale for commenting on the previous publication clearly described?**

Yes

**Are any opinions stated well-argued, clear and cogent?**

Yes

**Are arguments sufficiently supported by evidence from the published literature or by new data and results?**

No

**Is the conclusion balanced and justified on the basis of the presented arguments?**

No

**Competing Interests:** No competing interests were disclosed.

**I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

Reviewer Report 27 October 2015

<https://doi.org/10.5256/f1000research.7373.r9708>

© 2015 McCarthy M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



## Michael McCarthy

School of Botany, University of Melbourne, Parkville, Victoria, Australia

First, I apologise for the delay in writing this review – I’ve had other (also late!) reviews to conduct for other journals.

This article appears to be technically correct (e.g., the calculations in the box), but I think it makes some incorrect claims, and some other claims are either vague or unsubstantiated. I provide details below.

1. The authors write that Efron “concludes that an uninformative prior does not have the desired neutral effects on the output of Bayes’ rule”. Efron does not state this conclusion explicitly. My reading of Efron here is that he points out that the choice of the prior matters, and that using an uninformative prior can mislead. However, Efron does claim that Bayesian analyses based on uninformative priors are unreliable to the extent that they need to be checked with frequentist methods. Efron’s article undermines this point with the twins example because it is unclear how a frequentist analysis could be used to calculate the probability that the twins were identical. In fact, it seems impossible for a frequentist analysis to do that. This suggests that it is sometimes impossible to check an analysis with frequentist methods – to my mind this is the main problem with Efron’s position. In my opinion, the authors’ critique of Efron misses the mark.
2. The authors write “We argue that this example is relatively useless in illustrating Bayesian data analysis”. This seems an unfair statement. Efron’s example is useful for showing how prior information can influence the result, and that it is important to get the prior “correct”. This was the point that Efron was illustrating, and it seems reasonable - I disagree that the example is “relatively useless”.
3. The authors justify their claim that the “example is relatively useless” by writing “One reason is that Efron considers the particular set of twin boys as the entire population. In this case, statistics is not needed because there is no random sample drawn from a larger population.” This is a distraction from the main point. Efron was focused on this set of twin boys – it was an entirely appropriate use of Bayesian methods. Statistics is not only limited to inference about large populations.
4. The authors write “Bayes’ theorem is a mathematical equation, not a statistical model describing how data may be produced.” The model about how data are produced is summarised by the likelihood function in Bayes’ theorem. Therefore, it could be argued that Bayes’ theorem does include a model of how the data are produced.
5. The authors write “Efron uses this equation to show that the value on the left side of the equation changes when a term on the right side is changed, which is trivial and could be shown with any mathematical equation also in a non-Bayesian context.” This seems to miss the point. Efron shows that the posterior is sensitive to the choice of the prior. This seems reasonable given the audience,

even if it is already well known to those familiar with Bayesian methods. It seems unnecessary to criticise this aspect.

6. The authors write “Efron’s example can be rearranged so that it fits a more realistic situation in statistical data analysis”. It is unclear in what sense the authors’ example is more realistic. Efron’s twins example is drawn from a real-life query from friends - that seems “realistic” to me. In contrast, the authors’ example compares a flat prior and a strongly informative prior. The informative prior is close to being specified as a constant – the particular parameters for the prior were chosen arbitrarily because information about the degree of certitude was not available. Basing the analysis on arbitrary values does not give the impression of being more “realistic” than Efron’s example, and rarely would an informative prior be so precisely defined yet still be the subject of estimation with Bayesian methods. Further, defining informative priors with arbitrarily-chosen parameters does not seem to be best practice for Bayesian analysis. Overall, the author’s example does not seem ideal to illustrate their point.
7. The authors write that their “approach would be fairer than the calculations given by Efron”. The meaning of “fairer” is unclear. Both the authors’ approach and that of Efron show that the choice of prior influences the results. Why is one fairer than another?

**Is the rationale for commenting on the previous publication clearly described?**

No

**Are any opinions stated well-argued, clear and cogent?**

No

**Are arguments sufficiently supported by evidence from the published literature or by new data and results?**

No

**Is the conclusion balanced and justified on the basis of the presented arguments?**

No

**Competing Interests:** No competing interests were disclosed.

**I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

---

**Version 1**

Reviewer Report 24 December 2013

<https://doi.org/10.5256/f1000research.3175.r2816>

© 2013 McCarthy M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Michael McCarthy**

School of Botany, University of Melbourne, Parkville, Victoria, Australia

This paper by Amrhein *et al.* criticizes a paper by Bradley Efron that discusses Bayesian statistics (Efron, 2013a), focusing on a particular example that was also discussed in Efron (2013b). The example concerns a woman who is carrying twins, both male (as determined by sonogram and we ignore the possibility that gender has been observed incorrectly). The parents-to-be ask Efron to tell them the probability that the twins are identical.

This is my first open review, so I'm not sure of the protocol. But given that there appears to be errors in both Efron (2013b) and the paper under review, I am sorry to say that my review might actually be longer than the article by Efron (2013a), the primary focus of the critique, and the critique itself. I apologize in advance for this. To start, I will outline the problem being discussed for the sake of readers.

This problem has various parameters of interest. The primary parameter is the genetic composition of the twins in the mother's womb. Are they identical (which I describe as the state  $x = 1$ ) or fraternal twins ( $x = 0$ )? Let  $y$  be the data, with  $y = 1$  to indicate the twins are the same gender. Finally, we wish to obtain  $\Pr(x = 1 | y = 1)$ , the probability the twins are identical given they are the same gender<sup>1</sup>. Bayes' rule gives us an expression for this:

$$\Pr(x = 1 | y = 1) = \Pr(x=1) \Pr(y = 1 | x = 1) / \{\Pr(x=1) \Pr(y = 1 | x = 1) + \Pr(x=0) \Pr(y = 1 | x = 0)\}$$

Now we know that  $\Pr(y = 1 | x = 1) = 1$ ; twins must be the same gender if they are identical. Further,  $\Pr(y = 1 | x = 0) = 1/2$ ; if twins are not identical, the probability of them being the same gender is 1/2.

Finally,  $\Pr(x = 1)$  is the prior probability that the twins are identical. The bone of contention in the Efron papers and the critique by Amrhein *et al.* revolves around how this prior is treated. One can think of  $\Pr(x = 1)$  as the population-level proportion of twins that are identical for a mother like the one being considered.

However, if we ignore other forms of twins that are extremely rare (equivalent to ignoring coins finishing on their edges when flipping them), one incontrovertible fact is that  $\Pr(x = 0) = 1 - \Pr(x = 1)$ ; the probability that the twins are fraternal is the complement of the probability that they are identical.

The above values and expressions for  $\Pr(y = 1 | x = 1)$ ,  $\Pr(y = 1 | x = 0)$ , and  $\Pr(x = 0)$  leads to a simpler expression for the probability that we seek - the probability that the twins are identical given they have the same gender:

$$\Pr(x = 1 | y = 1) = 2 \Pr(x=1) / [1 + \Pr(x=1)] \quad (1)$$

We see that the answer depends on the prior probability that the twins are identical,  $\Pr(x=1)$ . The paper by Amrhein *et al.* points out that this is a mathematical fact. For example, if identical twins were impossible ( $\Pr(x = 1) = 0$ ), then  $\Pr(x = 1 | y = 1) = 0$ . Similarly, if all twins were identical ( $\Pr(x = 1) = 1$ ), then  $\Pr(x = 1 | y = 1) = 1$ . The "true" prior lies somewhere in between. Apparently, the doctor knows that one third of twins are identical<sup>2</sup>. Therefore, if we assume  $\Pr(x = 1) = 1/3$ , then  $\Pr(x = 1 | y = 1) = 1/2$ .

Now, what would happen if we didn't have the doctor's knowledge? Laplace's "Principle of Insufficient Reason" would suggest that we give equal prior probability to all possibilities, so  $\Pr(x = 1) = 1/2$  and  $\Pr(x = 1 | y = 1) = 2/3$ , an answer different from 1/2 that was obtained when using the doctor's prior of 1/3.

Efron (2013a) highlights this sensitivity to the prior, representing someone who defines an uninformative prior as a “violator”, with Laplace as the “prime violator”. In contrast, Amrhein *et al.* correctly points out that the difference in the posterior probabilities is merely a consequence of mathematical logic. No one is violating logic – they are merely expressing ignorance by specifying equal probabilities to all states of nature. Whether this is philosophically valid is debatable (Colyvan 2008), but this example does not lend much weight to that question, and it is well beyond the scope of this review. But setting  $\Pr(x = 1) = 1/2$  is not a violation; it is merely an assumption with consequences (and one that in hindsight might be incorrect<sup>2</sup>).

Alternatively, if we don't know  $\Pr(x = 1)$ , we could describe that probability by its own probability distribution. Now the problem has two aspects that are uncertain. We don't know the true state  $x$ , and we don't know the prior (except in the case where we use the doctor's knowledge that  $\Pr(x = 1) = 1/3$ ). Uncertainty in the state of  $x$  refers to uncertainty about this particular set of twins. In contrast, uncertainty in  $\Pr(x = 1)$  reflects uncertainty in the population-level frequency of identical twins. A key point is that the state of one particular set of twins is a different parameter from the frequency of occurrence of identical twins in the population.

Without knowledge about  $\Pr(x = 1)$ , we might use  $\Pr(x = 1) \sim \text{dunif}(0, 1)$ , which is consistent with Laplace. Alternatively, Efron (2013b) notes another alternative for an uninformative prior:  $\Pr(x = 1) \sim \text{dbeta}(0.5, 0.5)$ , which is the Jeffreys prior for a probability.

Here I disagree with Amrhein *et al.*; I think they are confusing the two uncertain parameters. Amrhein *et al.* state:

*“We argue that this example is not only flawed, but useless in illustrating Bayesian data analysis because it does not rely on any data. Although there is one data point (a couple is due to be parents of twin boys, and the twins are fraternal), Efron does not use it to update prior knowledge. Instead, Efron combines different pieces of expert knowledge from the doctor and genetics using Bayes' theorem.”*

This claim might be correct when describing uncertainty in the population-level frequency of identical twins. The data about the twin boys is not useful by itself for this purpose – they are a biased sample (the data have come to light because their gender is the same; they are not a random sample of twins). Further, a sample of size one, especially if biased, is not a firm basis for inference about a population parameter. While the data are biased, the claim by Amrhein *et al.* that there are no data is incorrect.

However, the data point (the twins have the same gender) is entirely relevant to the question about the state of this particular set of twins. And it does update the prior. This updating of the prior is given by equation (1) above. The doctor's prior probability that the twins are identical (1/3) becomes the posterior probability (1/2) when using information that the twins are the same gender. The prior is clearly updated with  $\Pr(x = 1 | y = 1) \neq \Pr(x = 1)$  in all but trivial cases; Amrhein *et al.*'s statement that I quoted above is incorrect in this regard.

This possible confusion between uncertainty about these twins and uncertainty about the population level frequency of identical twins is further suggested by Amrhein *et al.*'s statements:

*“Second, for the uninformative prior, Efron mentions erroneously that he used a uniform distribution between zero and one, which is clearly different from the value of 0.5 that was used. Third, we find it at least debatable whether a prior can be called an uninformative prior if it has a fixed value of 0.5 given without any measurement of uncertainty.”*

Note, if the prior for  $\Pr(x = 1)$  is specified as 0.5, or  $\text{dunif}(0,1)$ , or  $\text{dbeta}(0.5, 0.5)$ , the posterior probability that these twins are identical is  $2/3$  in all cases. Efron (2013b) says the different priors lead to different results, but this result is incorrect, and the correct answer ( $2/3$ ) is given in Efron (2013a)<sup>3</sup>. Nevertheless, a prior that specifies  $\Pr(x = 1) = 0.5$  does indicate uncertainty about whether this particular set of twins is identical (but certainty in the population level frequency of twins). And Efron's (2013a) result is consistent with  $\Pr(x = 1)$  having a uniform prior. Therefore, both claims in the quote above are incorrect.

It is probably easiest to show the (lack of) influence of the prior using MCMC sampling. Here is WinBUGS code for the case using  $\Pr(x = 1) = 0.5$ .

```
model
{
  pr_ident_twins <- 0.5 # prior probability that the twins are identical
  x ~ dbern(pr_ident_twins) # are they identical? If so, x = 1, and 0 otherwise

  pr_same_gender <- x + (1-x)*0.5 # the probability that the twins have the same gender. It equals 1 if x =
1, and 0.5 otherwise (i.e., if x = 0)

  same_gender <- 1 # the single data point - the twins are the same gender
  same_gender ~ dbern(pr_same_gender) # those data arise as a Bernoulli sample with probability
pr_same_gender
}
```

Running this model in WinBUGS shows that the posterior mean of  $x$  is  $2/3$ ; this is the posterior probability that  $x = 1$ .

Instead of using  $\text{pr\_ident\_twins} <- 0.5$ , we could set this probability as being uncertain and define  $\text{pr\_ident\_twins} \sim \text{dunif}(0,1)$ , or  $\text{pr\_ident\_twins} \sim \text{dbeta}(0.5,0.5)$ . In either case, the posterior mean value of  $x$  remains  $2/3$  (contrary to Efron 2013b, but in accord with the correction in Efron 2013a).

Note, however, that the value of the population level parameter  $\text{pr\_ident\_twins}$  is different in all three cases. In the first it remains unchanged at  $1/2$  where it was set. In the case where the prior distribution for  $\text{pr\_ident\_twins}$  is uniform or beta, the posterior distributions remain broad, but they differ depending on the prior (as they should – different priors lead to different posteriors<sup>4</sup>). However, given the biased sample size of 1, the posterior distribution for this particular parameter is likely to be misleading as an estimate of the population-level frequency of twins.

So why doesn't the choice of prior influence the posterior probability that these twins are identical? Well, for these three priors, the prior probability that any single set of twins is identical is  $1/2$  (this is essentially the mean of the prior distributions in these three cases).

If, instead, we set the prior as  $\text{dbeta}(1,2)$ , which has a mean of  $1/3$ , then the posterior probability that these twins are identical is  $1/2$ . This is the same result as if we had set  $\Pr(x = 1) = 1/3$ . In both these cases (choosing  $\text{dbeta}(1,2)$  or  $1/3$ ), the prior probability that a single set of twins is identical is  $1/3$ , so the posterior is the same ( $1/2$ ) given the data (the twins have the same gender).

Further, Amrhein *et al.* also seem to misunderstand the data. They note:

“Although there is one data point (a couple is due to be parents of twin boys, and the twins are fraternal)...”

This is incorrect. The parents simply know that the twins are both male. Whether they are fraternal is unknown (fraternal twins being the complement of identical twins) – that is the question the parents are asking. This error of interpretation makes the calculations in Box 1 and subsequent comments irrelevant.

Box 1 also implies Amrhein *et al.* are using the data to estimate the population frequency of identical twins rather than the state of this particular set of twins. This is different from the aim of Efron (2013a) and the stated question.

Efron suggests that Bayesian calculations should be checked with frequentist methods when priors are uncertain. However, this is a good example where this cannot be done easily, and Amrhein *et al.* are correct to point this out. In this case, we are interested in the probability that the hypothesis is true given the data (an inverse probability), not the probabilities that the observed data would be generated given particular hypotheses (frequentist probabilities). If one wants the inverse probability (the probability the twins are identical given they are the same gender), then Bayesian methods (and therefore a prior) are required. A logical answer simply requires that the prior is constructed logically. Whether that answer is “correct” will be, in most cases, only known in hindsight.

However, one possible way to analyse this example using frequentist methods would be to assess the likelihood of obtaining the data for each of the two hypothesis (the twins are identical or fraternal). The likelihood of the twins having the same gender under the hypothesis that they are identical is 1. The likelihood of the twins having the same gender under the hypothesis that they are fraternal is 0.5. Therefore, the weight of evidence in favour of identical twins is twice that of fraternal twins. Scaling these weights so they sum to one (Burnham and Anderson 2002), gives a weight of 2/3 for identical twins and 1/3 for fraternal twins. These scaled weights have the same numerical values as the posterior probabilities based on either a Laplace or Jeffreys prior. Thus, one might argue that the weight of evidence for each hypothesis when using frequentist methods is equivalent to the posterior probabilities derived from an uninformative prior. So, as a final aside in reference to Efron (2013a), if we are being “violators” when using a uniform prior, are we also being “violators” when using frequentist methods to weigh evidence? Regardless of the answer to this rhetorical question, “checking” the results with frequentist methods doesn't give any more insight than using uninformative priors (in this case). However, this analysis shows that the question can be analysed using frequentist methods; the single data point is not a problem for this. The claim in Amrhein *et al.* that a frequentist analysis “is impossible because there is only one data point, and frequentist methods generally cannot handle such situations” is not supported by this example.

In summary, the comment by Amrhein *et al.* raises some interesting points that seem worth discussing, but it makes important errors in analysis and interpretation, and misrepresents the results of Efron (2013a). This means the current version should not be approved.

## References

Burnham, K.P. & D.R. Anderson. 2002. Model Selection and Multi-model Inference: a Practical Information-theoretic Approach. Springer-Verlag, New York.

Colyvan, M. 2008. Is Probability the Only Coherent Approach to Uncertainty? *Risk Anal.* 28: 645-652.

Efron B. (2003a) Bayes' Theorem in the 21st Century. *Science* 340(6137): 1177-1178.

Efron B. (2013b) A 250-year argument: Belief, behavior, and the bootstrap. *Bull Amer. Math Soc.* 50: 129-146.

### Footnotes

1. The twins are both male. However, if the twins were both female, the statistical results would be the same, so I will simply use the data that the twins are the same gender.
2. In reality, the frequency of twins that are identical is likely to vary depending on many factors but we will accept 1/3 for now.
3. Efron (2013b) reports the posterior probability for these twins being identical as “a whopping 61.4% with a flat Laplace prior” but as 2/3 in Efron (2013a). The latter (I assume 2/3 is “even more whopping”!) is the correct answer, which I confirmed via email with Professor Efron. Therefore, Efron (2013b) incorrectly claims the posterior probability is sensitive to the choice between a Jeffreys or Laplace uninformative prior.
4. When the data are very informative relative to the different priors, the posteriors will be similar, although not identical.

**Is the rationale for commenting on the previous publication clearly described?**

No

**Are any opinions stated well-argued, clear and cogent?**

No

**Are arguments sufficiently supported by evidence from the published literature or by new data and results?**

No

**Is the conclusion balanced and justified on the basis of the presented arguments?**

No

**Competing Interests:** No competing interests were disclosed.

**I have read this submission. I believe that I have an appropriate level of expertise to state that I do not consider it to be of an acceptable scientific standard, for reasons outlined above.**

Author Response 27 Jul 2015

**Valentin Amrhein**, University of Basel, 4051 Basel, Switzerland

**We would like to sincerely thank Michael McCarthy for his thorough review, and we revised our paper accordingly. McCarthy's main point is that Efron's calculations and our approach differ because Efron's calculations are about one particular set of twins boys, while our analysis aims at making inference on the frequency of occurrence of identical twins in the larger population of twin boys. Indeed, we did not present a re-analysis of Efron's calculations but instead we used a data point that Efron casually cited, namely**

**that the twin boys turned out to be fraternal. Our aim was to show that a Bayesian data analysis is not the same thing as solving a mathematical equation such as Bayes' theorem. In our manuscript, we now clarified that our approach is different from the calculations provided by Efron. We also shortened the manuscript and removed statements that were criticized by Michael McCarthy.**

**Some other responses to McCarthy:**

"This claim might be correct when describing uncertainty in the population-level frequency of identical twins. The data about the twin boys is not useful by itself for this purpose – they are a biased sample (the data have come to light because their gender is the same; they are not a random sample of twins)."

**We agree that the sample would be biased if we were interested in the population of twins. If our population of interest are twin boys, however, the data are not biased.**

"Further, a sample of size one, especially if biased, is not a firm basis for inference about a population parameter."

**We agree this would be a low sample size if this were an empirical study. We only used the data point to illustrate how to update the knowledge about the probability of identical twins among twin boys.**

"Note, if the prior for  $\Pr(x = 1)$  is specified as 0.5, or  $\text{dunif}(0, 1)$ , or  $\text{dbeta}(0.5, 0.5)$ , the posterior probability that these twins are identical is  $2/3$  in all cases."

**In our view, the fixed posterior probability of  $2/3$  applies only to the prior specified as a fixed value of 0.5, while the other two prior distributions each produce posterior distributions of different shape. We thus would not agree to the notion that the posterior probabilities are identical in all cases (but we deleted the respective paragraph from our paper).**

**Competing Interests:** No competing interests were disclosed.

---

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact [research@f1000.com](mailto:research@f1000.com)

**F1000Research**