# Abel and Baker Redux: Probability and Description

Right at the beginning of the F&L readings for this week, you saw a well-known probability puzzler/"paradox." I'm going to risk angering the copyright gods and quote it in full here:

> Assume that boys and girls are born with equal frequency. Mr. Able says, "I have two children, and at least one of them is a boy." What is the probability that the other child is a boy? Mr. Baker says, "I went to the house of a two-child family, and a boy answered the door." What is the probability that the other child is a boy?

The readings then assert to you that the answer to Mr. Able's problem is 1/3, and the answer to Mr. Baker's problem is 1/2.

I'll bet you have trouble believing this. But it's correct. Here's one way to think about why: **probability statements are really sensitive to how you carve up the problem space**.

The Abel problem could be restated as follows: first, take all the possible families with two kids, and then eliminate those where both are girls. Then, just from the set of families remaining, what proportion of them have two boys?

The Baker problem could be restated as follows: first, take all the possible families with two kids. Of them, go knock on their doors, and, from those doors you knocked on in which a boy answers, in what proportion of them is the second kid a boy?

The catch is that the sets you're looking into are different. In the Abel problem, you're looking at *every family that has a boy* and asking how likely it is that those families have two boys. In the Baker problem, you're looking at *every family that has a boy AND where that boy happened to answer the door*, and asking how likely it is that those families have two boys.

You would expect the size of the set you're looking at in the Baker problem to be smaller. After all, there are some families that have both a boy and a girl for which the girl will happen to answer the door. Intuitively, this should suggest to you that the pool of families for which a boy has answered the door would be smaller and, importantly, more weighted toward families with more boys. After all, in a boy/girl family, there's only a 50% chance that a boy will answer the door, whereas in a boy/boy family it's guaranteed that a boy will answer. So constructing our set of families where a boy answers means we lose some of the boy/girl families but none of the boy/boy families (we can safely assume no parents answer in any event since parent answering would be independent of which gender of the child answers and what genders there are among children in the family). The pool will be relatively more full of boys.

## Now let's do the math.

Let's formalize the Baker problem with Bayes Rule. Reminder:

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}$$

What we want to know is the probability of both kids being a boy, conditional on the door being opened by a boy. So, in our equation above, assign both-kids-boy to B, and boy-opened-door to A.

And here's what we do know:

- Probability of door being opened by a boy, conditional on both kids being boys — $P(A|B) = 1$.
- Probability of both kids being boys — $P(B) = 1/4$ because the available packages of kids are $\{bg, gb, bb, gg\}$ and they have equal probability, and we know that as mutually exclusive and exhaustive events they must sum to 1.
- Probability of the door being opened by a boy, or $P(A)$. This one we have to do a little bit more work to construct. Let's reason it through. We know all of the available packages of kids (that is, events), which are mutually exclusive. And we know the probability of each of those packages appearing (1/4). We also know the probability of getting a boy to answer the door under each of the packages of kids. So we can just multiply those probabilities out and add them per the rules of probability we've already established. For example, the probability of bg is .25, and the probability of getting a boy answering the door there is 1/2, and so forth. We ultimately end up with $P(A) = .25 \cdot .5 + .25 \cdot .5 + .25 \cdot 1 + .25 \cdot 0 = .5$.

Now we can plug in:

$$P(B|A) = \frac{1 \cdot .25}{.5} = \frac{1}{2}$$

Just as the readings told us.

How about the Abel problem? Well, we don't even need Bayes Rule for this one, we just need to observe that 1/3 of the families with one boy have both boys. But let's do it anyway, just as an exercise. Set B to be, yet again, "family has both boys," and set A to be "family has one boy."

- Probability of a family having one boy, conditional on their having both boys — $P(A|B) = 1$, obviously.
- Probability of a family having both boys — $P(B) = 1/4$ because, again, our available packages of kids are $\{bg, gb, bb, gg\}$, with equal probability, and their probabilities have to all add up to 1.

- Probability of a family having one boy — $P(A) = 3/4$ because of 4 available equal probability packages only one of which has no boys.

Again, plug that sucker into Bayes rule:

$$P(B|A) = \frac{1 \cdot .25}{.75} = \frac{1}{3}$$

Once again, the readings did not lie to us.

We should take several important lessons from forcing ourselves to work out this example all the way.

1. Clear talking facilitates clear thinking. With any probability statement, precise mathematical formulation is better than words, and more detailed words are better than less detailed words.
2. Probability statements are sensitive to the sample space (as the reading points out), or, more colloquially, the set of possible events being looked at. Make absolutely sure you have a clear idea what that sample space is, and whether it's the correct one or not.
3. Almost any supposed paradox that you see can be resolved with aggressive application of Bayes Rule. Bayes Rule is the closest thing we have to a magic wand for fixing screwed-up thinking about probability.
4. When in doubt about some probability statement, simulate it! You just did a ton of work learning basic coding, it's time to get some payoff from it. The great thing about simulations is that they force you to make your assumptions explicit by putting them in a form the computer can understand. If you can't define your sample space, you can't ask our good friend Python to simulate it. Often, writing out a simulation will force you to resolve ambiguities in how you structure the problem, and give you the boost you need to understand it yourself. And if you still aren't quite sure even after writing the simulation out, you can run it and see what happens!

I've written a simulation of the Able/Baker problem down below. Before you look at it, you might consider writing your own simulation. But if you don't have time, feel free to run my code with sufficiently large values and see what you get.

.

.

.

.

.

.

.

.

```
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
.
import random
def abel(num_families):
    fams = []
    for fam in range(num_families):
        kid1 = random.choice(["b", "g"])
        kid2 = random.choice(["b", "g"])
        if (kid1 == "b") or (kid2 == "b"):
            fams.append((kid1, kid2))
    two_boy_fams = [x for x in fams if x == ("b", "b")]
    return {"total fams with a boy": len(fams),
            "total fams with two boys": len(two_boy_fams)}

def baker(num_families, num_knocks):
    if num_knocks > num_families:
        raise ValueError("Can't knock on more doors than there are households.")
    fams = []
    for fam in range(num_families):
        kid1 = random.choice(["b", "g"])
        kid2 = random.choice(["b", "g"])
        fams.append((kid1, kid2))
    random.shuffle(fams)
    knocks_with_first_boy = []
    for knock in range(num_knocks):
        thisfam = fams.pop()
        kid_who_answers = random.choice([0, 1])
        if thisfam[kid_who_answers] == "b":
            otherkid = int(not kid_who_answers) # flips 0 to 1 and 1 to 0.  Actually don't
            knocks_with_first_boy.append(thisfam[otherkid])
    knocks_with_second_boy = [x for x in knocks_with_first_boy if x == "b"]
    return {"total knocks where a boy answered": len(knocks_with_first_boy),
            "knocks where boy answered and second kid is a boy": len(knocks_with_second_boy
```

As your reward for reading this far, I'll tell you a secret. I wasn't sure why the Able/Baker problem came out the way it does either—not until I wrote the simulation code. But, just as I said, writing this code forced me to think about what the sample space for the Baker problem actually is, and hence guided me to the rest of this explanation. Even those of us with years and years of probability experience need a little help sometimes.