# Switch to Win?

The Monty hall problem is an extremely simple, yet counter-intuitive probability puzzle that baffled even the most brilliant minds back in the late 1900s. The scenario is based loosely off an American game show called "Let's Make a Deal" and named after the original host, Monty Hall. The premise of the puzzle is as follows: You are a contestant on a game show, and in front of you are 3 doors. Behind 2 of them are goats and behind 1 of them is a brand new (insert your dream car here). The host knows which door holds the new car and asks you to pick a door first. Let's say you pick door 1, the host then opens door 2, reveals a goat and says to you, "Do you want to pick door 3 instead?" The question is then, whether or not switching doors will give you a higher chance of winning a car.

INSERT PICTURE HERE

The most obvious and intuitive answer would be no. This is because, regardless

of what the host tells you, the chance of you picking the right door is still 1/3 or 33%, isn't it? After all, he may just be out to trick you into switching from a car to a goat! Well, unfortunately, there is a little more to this problem than meets the eye. In the following sections, we will prove why switching is always better in 3 ways:

* Bayesian Statistics
* Intuitive Logic
* Computer Generated Simulations

As well as to go through the underlying misconceptions and how it may apply to other problems.

## Intuition

When faced with a problem that is fairly mathematical in nature, there are a variety of methods to go about solving it. One of the ways that work here is to "*take things to the extreme*". Although there is no formula for when or how to use this technique, it is generally used to give the person a different perspective on the problem with the hopes of finding some new insights (maybe), or to test the robustness of a theory to see if it makes logical sense throughout the whole spectrum of possibilities. Back to the Monty hall problem, the key thing to note here is that the host, actually knows which door the car is behind. Although it may seem unrelated for now, let us "take it to one extreme" where instead of 3 doors, there are now 100 doors with 99 of them holding goats and only 1 of them with your dream car. Let's say now you pick door 1 randomly, and the hosts then opens 98 doors (door number 2 to 99) and reveals goats behind them. He then asks you again if you would want to switch to door 100. How confident are you on your initial choice now? Without going into anything technical yet, once we actually see the problem in this perspective, I think it is safe to say that most people would probably switch doors because *the host knew which door the car was in!*

Now that we have established intuitively that it is better to switch with 100 doors, we have a starting point, and the next step is to see if this behavior actually extends to case with only 3 doors.

## Bayesian Statistics

The algebraic proof in itself is a simple one but the underlying concept relates to Bayesian statistics, which, to the majority of people (or students) would be the formula:

Before we delve into how this works, we have to establish some concepts:

1. P(B) just means the probability of B occurring and P(B') is the probability of B **not** occurring or 1 - P(B).
2. P(B|A) is the probability that B occurs given that A has occurred.
3. P(B|A) \* P(A) = P(BA) This is because, the probability of A occurring first, then B occurring given A has occurred is the same as the probability of B and A occurring. (only if B and A are independent events, example: successive coin tosses)
4. P(B) = P(BA) + P(BA') The probability of B occurring can be broken down into the sum of: (B occurring, A occurring) + (B occurring, A not occurring)
5. P(AB) = P(BA) Does not matter which occurs first for independent events (example tossing which coin first does not matter) These underlying concepts can be better understood with the help of a tree diagram:

INSERT PICTURE HERE

NOTE: At first glance it may seem weird that we want to find the

probability of an earlier event (A) given that a later one (B) has happened, but this is just because there is a lack of context in this explanation. Just as an example, imagine A was the weather (sunny or rain) and B was the activity (sports or sleep). It would seem natural to calculate P(B|A), example: The probability I would play sports given some weather. However, it is also perfectly normal to want to calculate the chances of where someone is using P(A|B) (whether a person is out or at home) given the current weather! Okay, so now using the formula in tandem with the tree diagram, we can break down the right side of the formula into the numerator and denominator.

INSERT PICTURE HERE

Let's first talk about the denominator. Intuitively, if we are finding P(A|B) the denominator must be all the events where B has already occurred. This is represented by all the scenarios where the lines are dashed. This includes the 2 components, P(B|A') \* P(A') + P(B|A) \* P(A). From point 3), this can be reduced to P(BA') + P(BA) and finally, using point 4) we can see that it is just P(B), which makes perfect sense! Secondly, the numerator is given as P(B|A) \* P(A), which is just P(BA) = P(AB) from the derivation just above. This is then represented by the red dashed line in the diagram above. So finally, putting it simply, P(A|B) is the probability of both A and B occurring, under the condition that only B has occurred, which is just the intersection (A n B) divided by the blue shaded area below instead of the whole box.

INSERT PICTURE HERE

Okay, so back to the Monty hall problem. For simplicity, let’s have:

* D1: The Event of the host opening door 1.
* D2: The Event of the host opening door 2.
* D3: The Event of host opening door 3.
* C1: The car is behind door 1.
* C2: The car is behind door 2.
* C3: The car is behind door 3.

Initially, we can safely say that P(C1) = P(C2) = P(C3) = 1/3. This is because we do not know the position of the car and would just randomly choose our first door. Assuming we choose door 1 and the car **is** in door 1, the host can choose to open either door 2 or 3 to reveal a goat. This means that P(D3|C1) = P(D2|C1) = 1/2. However, if the car is in door 2, then the host would have to open door 3 to reveal the goat and P(D3|C2) = 1. Also, the host will not reveal the door with the car, so we can safely assume that P(DX|CX) = 0 for X = 1,2 and 3. Now let us further assume that the host opens door 3. Using Bayes’ formula:

Here, P(D3) = P(D3C1) + P(D3C2) + P(D3C3) = P(D3|C1) \* P(C1) ... based on points 3) and 4) above. This above result shows that, if I had picked door 1 initially and the host opens door 3, the probability that the car is in door 2 is double the probability that it is in door 1, my initial selection! So, in this case switching does increase (doubles actually) my chances of getting my dream car! This formula can be extended to different scenarios and all would still yield the same result! For those who hate interpreting formulae and expressions, the last method shown eliminates the need for all that, simulations.

## Simulations

With the current level of computing power, anyone and everyone with a laptop with Microsoft Excel installed can recreate the scenario with a few simple formulae and run it many times to get the behavior and patterns of decisions made. This is also called Monte-Carlo simulations and I’ll leave the concepts and reasoning behind it for another time. If you are interested in the code, there is a link to by GitHub [here](https://github.com/jtsw1990/monty_hall_simulations). After running different numbers of simulations from n=1 all the way to n=1000, I plotted the resulting win rates in the graph below.

INSERT GRAPH HERE

As you can see, although there is a little overlap when the number of simulations are low due to pure randomness, the law of averages will start to kick in as the number of simulations increase and the win rates for the switching strategy and the non-switching strategy will slowly converge to approximately 66% and 33% respectively. This not surprisingly also corresponds to our mathematical derivation where switching doors will double your chances of winning!

## So what good is this?

Sadly, knowing this now would not win you a car because during this day and age, our game shows have evolved to be much more complicated and pointless at the same time (a similar one would be an old game show called Deal or No Deal). However, important things to takeaway would be some problem-solving techniques, a (hopefully gentle) introduction to Bayesian statistics and some coding techniques for those interested. Remember, not everything that is intuitive will be correct, and not everything that is correct will be intuitive, so as they say, you better check yourself before you wreck yourself.