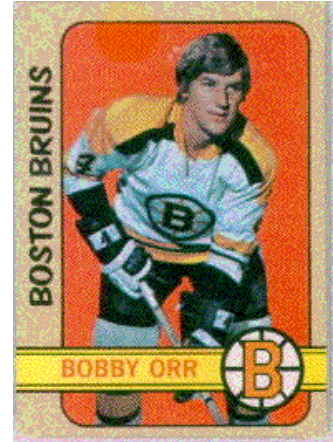


# Collector's problem

Byron Schmuland  
 Department of Mathematical Sciences  
 University of Alberta, Edmonton  
 schmu@stat.ualberta.ca

If you've ever collected sports cards, you know what it's like. At first, every pack you buy adds to your rapidly growing collection. But pretty soon you are collecting a lot of duplicates, and in a short time your duplicate pile is bigger than your collection. Eventually, getting a new card for your collection is a rare event, most of the time the whole pack goes straight into the duplicate pile. Finally, you need only one last card, the elusive Bobby Orr, and you spend several weeks buying pack after pack before you finally finish off your collection.



Are you being ripped off? Does it make sense that your duplicate pile should end up being two or three times bigger than your collection? Is the card company artificially creating rare cards to increase sales?

Let's try to understand the mathematics of the collector's problem. For simplicity, let's assume that you buy cards one at a time, and that every player's card has an equal chance of turning up. This says that the card company is not cheating: there are as many Bobby Orr's as there are Bobby Schmutz's. We let  $n$  be total number of different cards. When I collected hockey cards,  $n$  was about 250. This is hard to handle so let's warm up with a simpler problem where  $n = 4$ .

Start with a well-shuffled deck of cards. Randomly pull out a card, replace and repeat. How long does it take, on average, to get all four suits? Try it a few times, and you'll see that you very rarely get all four suits with the first four cards, usually it will take eight or nine, and occasionally over a dozen cards before all four suits show up. Here are some sample results:

7 cards:	↓   ↓ ↓   ↓ ♠ ♠ ♣ ♦ ♣ ♦ ♥ ♥ ♣ ...
5 cards:	↓ ↓   ↓ ↓ ♣ ♥ ♥ ♦ ♠ ♦ ...
14 cards:	↓   ↓   ↓   ↓ ♣ ♣ ♥ ♥ ♣ ♥ ♣ ♣ ♠ ♣ ♣ ♠ ♥ ♦ ♣ ...

I've marked each new suit with a little arrow, and the fourth new suit means we're done. To do the mathematics, it is convenient to divide the total into the four pieces

between new suits. Let's call

$$E(T_1) = 1$$

$E(T_2)$  = the average number of cards between the 1<sup>st</sup> and 2<sup>nd</sup> new suit


$E(T_3)$  = the average number of cards between the 2<sup>nd</sup> and 3<sup>rd</sup> new suit

$E(T_4)$  = the average number of cards between the 3<sup>rd</sup> and 4<sup>th</sup> new suit.

The average number of cards to get all four suits is  $E(T_1) + E(T_2) + E(T_3) + E(T_4)$ .

### Flip it over!

If the chance of a random outcome is  $p$ , then on average you need  $1/p$  trials until this outcome occurs.

For instance, if you toss a fair coin, you need 2 throws on average to get a head, and with a fair die you need 6 throws on average to get a . I guess this makes sense. The rarer the outcome, the longer it takes to see it.

Back to the card problem. The first equation  $E(T_1) = 1$  is easy, since the first card always gives a new suit. Now after you have one suit, the chance of a new suit is  $3/4$  and the average number of cards until this happens is (flip it over!)  $E(T_2) = 4/3$ . After you have two suits, the chance of a new suit is  $1/2$  and the average number of cards until this happens is  $E(T_3) = 2$ . Finally, after you have three suits, the chance of a new suit is  $1/4$ , so  $E(T_4) = 4$ . This gives the final answer of

$$E(T_1) + E(T_2) + E(T_3) + E(T_4) = 4 \left( \frac{1}{4} + \frac{1}{3} + \frac{1}{2} + \frac{1}{1} \right) = 8 \frac{1}{3}.$$

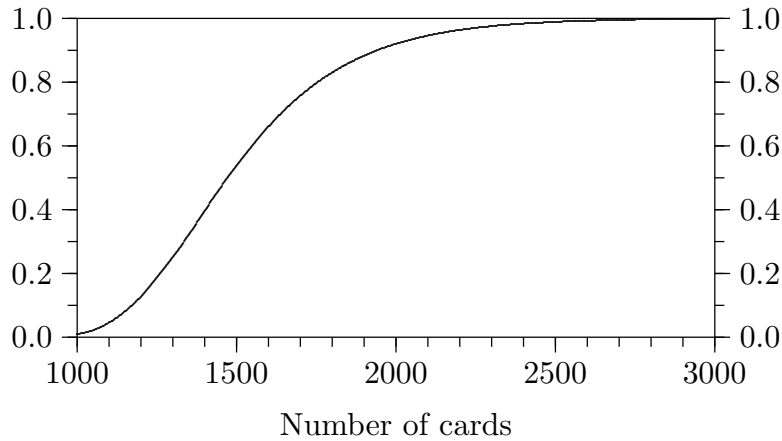
The nice thing is that the same pattern works no matter how big the problem is. Going back to hockey cards, if there are a total of  $n$  cards to be collected, then the average number of random selections until all  $n$  different cards appear is

$$n \left( \frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + \frac{1}{1} \right) \approx n(\log(n) + .577),$$

where  $\log(n)$  means the natural log of  $n$ . Plugging in  $n = 250$ , I find that the average number of cards purchased to get a full collection is about 1525. By the time you're done, the average duplicate pile should be five times larger than the collection. Well, that explains a lot!

However, *average* doesn't mean *typical*. Some hockey card collectors will need to buy 1525 cards or so, but some lucky people will get away with less, and some unlucky ones will need more. Further analysis of the collector's problem gives us the following chart.

Probability to complete a collection of size  $n = 250$



From this picture we can see that virtually everybody will have to buy between 1000 to 2500 cards to complete their collection. So the card companies don't need to deliberately create shortages. Just like for casino owners, random chance alone guarantees them big sales.

So what's a poor hockey card collector to do? The answer: Trade with your friends! Some of the cards you want are in your friend's duplicate pile and vice versa. The mathematics is more difficult, but if two collectors cooperate, then the average number of purchases to obtain two complete collections is about  $n(\log(n) + \log \log(n) + .577)$ . For  $n = 250$ , this means that you and your friend can expect to buy about 1952 cards before getting two full collections. This is about 976 purchases per person, much lower than the 1525 when you go it alone. In fact, the more friends you have trading, the cheaper you *all* get your complete collections.

Probability to complete two collections of size  $n = 250$

