THE JAMES WILLOUGHBY COLUMN

THE SIMULATION GAME

Last month, the World's Best Racehorse Ratings crowned the US superstar Flightline as the equal of Frankel on a rating of 140. This marks two super horses born within a decade of each other. What makes these horses so great athletically? How are they so much better than any other horse around? Sometimes, the answer to complex questions like these is to simulate.

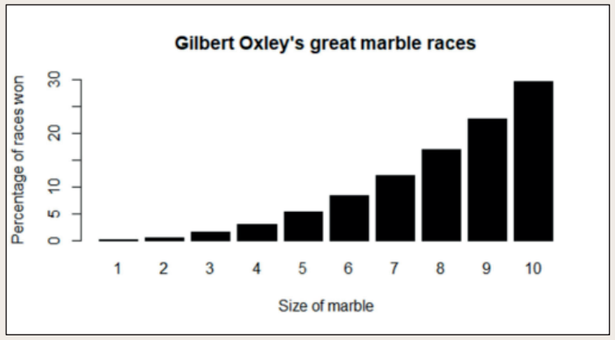
IN THE final year of his life, Gilbert Oxley lost his memory. But he never lost his marbles.

On the top of a sideboard in the back room of his house, my grandad Gilbert kept a mottled green, vinyl-clad box with a silver clasp. Resting on black cloth inside were 10 glass spheres, each flecked by a distinct combination of colours, each a different size and smoothness.

These marbles were his stable of racehorses, and together we would simulate races by allowing them to roll down his tarmac drive and collect the results. There was a point to this beyond entertainment, however, which came from my grandad's curiosity about how racing worked. "What I cannot create, I do not understand," the great physicist Richard Feynman said, of trying to solve problems this way.

My grandad thought great horses had characteristics which multiplied the effect of each other. A long stride is an obvious advantage to a runner, for instance, but stride-length was just one of a suite of characteristics associated with size, and bigger horses might also tend to have bigger hearts and bigger lungs which, in addition to the ground they covered, could keep them going longer.

The marbles which were to show how this worked had 10 different diameters. Each grade of marble was just the same step-size bigger than the next. As they rolled down the slope, they would collide with both each other and tiny bits of tarmac. The effect was like Escalado, if you know that horse racing game.



Size	1	2	3	4	5	6	7	8	9	10
Win %	0.1	0.5	1.4	3.0	5.2	8.3	12.1	17.0	22.7	29.7
Increase		*0.4*	*0.9*	*1.6*	*2.2*	*3.1*	*3.8*	*4.9*	*5.7*	*7.0*

Figure 1: the results of simulating races by rolling marbles down a tarmac drive in 1983

After many races our results would look something what is shown in figure 1 below.

What is causing the results in Figure 1?

As the diameter of a marble increases, the percentage of races it wins increases.

Size is increasing by the same step, but strike rate is going up by an increasing amount. As the table under the graph shows, grade 2 marbles won 0.5% of all races which is only 0.4% more than grade 1 marbles. Further up the scale, however, grade 10 marbles – the biggest in diameter – won 7% more races than grade 9 ones, a huge difference.

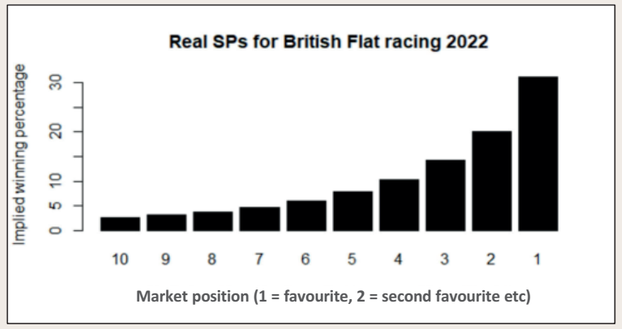
It turned out that the size of a marble was not the only characteristic that got it to the winning line faster. Smoothness was also important because perfectly spherical marbles roll straighter than those with imperfect curvature. Bigger marbles also happened to be smoother, probably because of the way they were made.

Because size and smoothness were correlated in the marble population, a multiplicative effect was at work between the two characteristics. As a marble rolled and survived collisions with competitors and debris, bigger marbles had both greater velocity owing to their smoothness and greater momentum due to their mass; both quantities are important in getting to the winning line first.

The analogue here is that the size of a marble is a horse's stride-length and the smoothness of a marble its action. If grade 20 marbles had existed, we might have been looking at the Frankel or Flightline of marble races. (Or the Brigadier Gerard, as my grandad would have said at the time.)

Is the marble simulation realistic?

The question is: does real racing follow the same pattern as marble races? Does the gulf in class between horses widen or narrow as you ascend the class scale? It might seem that ratings are one way to answer these questions, but I have shown many times that ratings like those produced by the BHA to set handicap weights are biased: top weights win more often than bottom weights.



Fav	10	9	8	7	6	5	4	3	2	1
Win %	2.6	3.1	3.7	4.7	6.0	7.8	10.3	14.2	20.0	31.1
Increase		*0.5*	*0.6*	*1.0*	*1.3*	*1.8*	*2.5*	*3.9*	*5.8*	*11.1*

Figure 2: winning percentages implied by SPs have the same distribution as marble races

As researchers Peter Bebbington and Julius Bonart realised in 2016, betting markets are an excellent gauge to the rational expectations of horses. Figure 2 uses the Flat results for last year in Britain to show the average win percentages implied by the starting price of a horse according to its market position, with 1 being the favourite, 2 the second favourite etc.

Figure 2 shows that the distribution of talent for real horse races mirrors that for marbles. Notice that when the runners are arranged in market order, the height of the bars representing implied winning percentage is similar to that of winning percentage in our marble races. (A technical note: bars on the left of Figure 2 are taller than in Figure 1 because SP markets include the Favourite-Longshot bias – horses with only an outside chance tend to be relatively over bet which unreasonably inflates their implied winning percentages.)

What can we learn from the marble simulation of races?

A mathematical or computer simulation of a real system is knowingly simplified compared with real life. This allows its most influential components to emerge from the soup of factors which make it work. My grandad noticed that bigger marbles also tend to be smoother, and the combination of both factors makes them exponentially better than the rest.

It is a reasonable presumption that something analogous to this process produced Frankel and Flightline. These horses had longer strides but were also faster-twitch than ordinary equine athletes. Positive attributes are not just additive but multiplicative in the athletic arena, as we have seen in the marble world.

The randomly broken stick

Bebbington and Bonart described the effect of Table 2 as akin to the length of pieces when a stick is randomly broken into parts. It is difficult to explain why that is an idea driven by the mathematical process, but suffice to say they found the same pattern in betting markets was no fluke. They concluded: 'The ability of a horse can be defined in such a way that its winning probability is the ratio of its ability to the sum of all its competitors' abilities – provided ability is exponentially distributed.'

This has profound implications for how horses should be rated and how their athletic ability is deconstructed. Traits multiply, not add, so rating horses on a linear scale between 0 and 140 is hopeless, if the predictions for future races depend on the one point equals one pound of weight.

Racehorse abilities are separated by magnitudes, like the destructive power of earthquakes.

Mendel was wrong, and so was Tesio

That relative racehorse talent should be described by the randomly broken stick is also interesting from a genetic standpoint. In the 19th century, the father of modern genetics Gregor Mendel proposed his Law of Independent Assortment, the upshot of which was that characteristics were inherited independently of one another. This was presumably still the prevailing wisdom when parroted by the brilliant Federico Tesio in his seminal work *Breeding The Racehorse.*

'I had thought originally that exactly the opposite was the case,' Tesio wrote. 'That each inherited character was invariably accompanied by two or three others in a standard combination… but each characteristic will be inherited [independently] according to the law of Mendel.'

Later research showed that genes that are close together in the chromosome are more likely to be inherited in pairs than two randomly situated genes. So, it makes sense to me that if you breed generations of the same animal and select the best for breeding, genes which are correlated with ability tend to co-exist more frequently than by chance. If races between horses have something in common with races between marbles, then it could easily be inferred that this artificial selection makes it a little more likely that longer-striding horses are also better movers, say.

Statistically, Frankel and Flightline were rare horses, but the incidence of two such animals within 10 years is no more than interesting, rather than statistically significant. It is not possible to hack the underlying process which resulted in their appearance, but at least simulation of a simpler analogue to the system can enable some understanding of how such utterly superior horses dominate. ■