THE JAMES WILLOUGHBY COLUMN

This month James discusses data on links in breeding between the sire and the dam's sire, and evaluates 'nicks' -- the combination of certain bloodlines that produces an unusually high probability of a successful foal.

EVERYONE INTERESTED in racehorse breeding understands that both the sire and the dam contribute to the genetic potential of a foal. Yet, when it comes to assessing sires, little effort is made in statistical tables to allow for the average quality of the mares who were covered.

Horse	Sire	Damsire	Sire Index	Damsire Index	Combined index	Prob	Odds
City Of Troy	Justify	Galileo	1.87	1.37	1.87 x 1.37 = 2.56	0.162	5.2-1
Alyanaabi	Too Darn Hot	Kitten's Joy	1.40	1.15	1.4 x 1.15 = 1.61	0.102	8.8-1
Eben Shaddad	Calyx	Galileo	1.42	1.37	1.42 x 1.37 = 1.95	0.123	7.1-1
Henry Adams	No Nay Never	Galileo	1.56	1.37	1.56 x 1.37 = 2.14	0.135	6.4-1
Haatem	Phoenix Of Spain	Cape Cross	1.50	1.24	1.5 x 1.24 = 1.86	0.117	7.5-1
Iberian	Lope De Vega	High Chaparral	1.73	1.20	1.73 x 1.2 = 2.08	0.131	6.6-1
Array	No Nay Never	Oasis Dream	1.56	1.23	1.56 x 1.23 = 1.92	0.121	7.3-1
Indian Run	Sioux Nation	Danehill Dancer	1.44	1.21	1.44 x 1.21 = 1.74	0.110	8.1-1

Figure 1: odds for the Dewhurst Stakes at Newmarket based only on each runner's sire and damsire combined

The truth is that this is not straightforward. In the US, indices such as the earnings of a mare on the racetrack are sometimes used as a proxy for quality, but some of the most successful broodmares were unraced or far less successful on the track than at stud.

Sire	Wins	Runs	SR	Index	Damsire	Wins	Runs	SR	Index
Frankel GB	704	3766	19	2.35	Sea The Stars IRE	250	1710	15	1.45
Dubawi IRE	1597	8405	19	2.28	Galileo IRE	2012	15057	13	1.37
Galileo IRE	1672	10647	16	2.25	Frankel GB	45	298	15	1.35
Deep Impact JPN	51	275	19	2.14	Street Cry IRE	474	3531	13	1.34
Sea The Stars IRE	839	4920	17	2.13	Dubawi IRE	666	4971	13	1.34
Kingman GB	547	3336	16	2.03	Shamardal USA	858	6468	13	1.34
Justify USA	27	130	21	1.87	Authorized IRE	174	1325	13	1.34
War Front USA	321	1991	16	1.87	Monsun GER	274	2037	13	1.32
New Bay GB	216	1246	17	1.86	Makfi GB	91	640	14	1.39
Night Of Thunder IRE	318	2122	15	1.83	Maria's Mon USA	108	812	13	1.32
Golden Horn GB	252	1674	15	1.78	Sixties Icon GB	55	377	15	1.32
Lope De Vega IRE	943	6640	14	1.73	Lope De Vega IRE	95	699	14	1.32
Harry Angel IRE	102	631	16	1.72	Le Havre IRE	47	328	14	1.32
Farhh GB	173	1163	15	1.71	Kingmambo USA	732	5475	13	1.31
Roaring Lion USA	69	434	16	1.70	Horse Chestnut SAF	26	144	18	1.31
Pride Of Dubai AUS	107	789	14	1.69	Poliglote GB	28	172	16	1.31
Le Havre IRE	183	1291	14	1.69	Hard Spun USA	69	510	14	1.31
Scat Daddy USA	124	834	15	1.66	Dream Ahead USA	66	468	14	1.31
Shamardal USA	1182	7993	15	1.66	Unbridled's Song USA	123	902	14	1.30
Monsun GER	104	685	15	1.64	Songandaprayer USA	34	223	15	1.30
Wootton Bassett GB	121	875	14	1.63	Marchand De Sable USA	45	309	15	1.30
Havana Grey GB	169	1134	15	1.62	Whipper USA	223	1761	13	1.30
American Pharoah USA	38	277	14	1.61	Keltos FR	26	164	16	1.30
Siyouni FR	333	2532	13	1.59	Toccet USA	26	168	15	1.30

Figure 2: indices of independent effects of leading sires and broodmare sires, Britain and Ireland 2010 - 2023

From a statistical standpoint, judging the quality of a mare's offspring presents the problem of the small sample. Mares have few foals, and it takes many years to assess their potency. The racetrack earnings of a mare are therefore a useful statistical proxy for a mare's quality, even if earnings on the track often do not translate to success in the sheds.

Another convenient way to estimate the average quality of a stallion's mates is the identity of the broodmare sire. If one mare is by Galileo and another is by a much lesser stallion, it is a lot more likely that the Galileo mare will produce quality racing stock.

If we have a database of results, it is a straightforward task statistically to find the independent contributions of sire and damsire using a technique called maximum likelihood.

Take this season's Group 1 Dewhurst Stakes at Newmarket. If we ignore each horse's form, and merely think of them as having a win potential coming from their sire and damsire combined, the odds of winning would look like as shown in Figure 1.

How does this work? First, the indices in columns 4 and 5 are based on the results of 103,839 races which have taken place in Britain and Ireland from 2010 to 2023. The Index is known as an Odds Ratio and is a smarter version of what is better known in racing as Impact Value. Odds Ratios use not just wins and losses like Impact Value but weight the wins and losses according to the strength of the opposition. So, Odds Ratios are Impact Values adjusted for class.

In the case of City Of Troy, we can read from the sixth columns of the table that the combination of the Odds Ratio (referred to from now on as just 'Index') of his sire (1.87) and damsire (1.37) indices is 2.56. In effect, this means that the offspring of Justify and a Galileo mare has won 2.56 times more often in the data than an average sire and damsire combination. Don't worry about the derivation, just remember that 1.00 is a cross between sire and damsire of average quality.

Odds Ratios are particularly useful when we know all the Odds Ratios in a race. If we use the letter j as a variable to represent each horse in a race (j=1, 2, 3, .., n where n is the field-size) then the probability of a win is simply:

$$\Pr(\text{win}) = \frac{Odds\ Ratio_j}{\sum Odds\ Ratio} \qquad [\ \textstyle\sum \text{ symbol is read 'sum of' }]$$

That is to say: a horse's win probability is just the proportion of its Odds Ratio to the sum of all Odds Ratios (including its own) in a race. These probabilities for the Dewhurst Stakes are shown in the penultimate column and the equivalent odds in the final column.

Of course, City Of Troy started at 15-8 on rather than 5-1 against because a lot more was known about him than him just being an average product of Justify and a Galileo mare. These other factors – known as covariates in statistics – are a combination of objective factors such as his form, class and potential which can be themselves evaluated in the Odds Ratio framework.

Who are the leading sires and broodmare sires?

It should not be a surprise that some familiar names appear in the table of leading sires and broodmare sires by the method of maximum likelihood contained in Figure 2.

Notice that the Odds Ratios of the leading broodmare sires are lower than those of the leading sires. This is to be expected because their influence is one generation removed. Some interesting names reside in the left panel, notably young stallions Justify, Harry Angel and Roaring Lion, though the last of these is sadly dead.

Some of the stallions in the left panel are long in the tooth or no longer around, so it may be more interesting to look at the Indices of first- and second-crop stallions whose stock appeared in 2023. These are listed in Figure 3.

Sire	Wins	Runs	SR	Index
Justify USA	27	130	21	1.87
Harry Angel IRE	102	631	16	1.72
Roaring Lion USA	69	434	16	1.70
Havana Grey GB	169	1134	15	1.62
Blue Point IRE	54	335	16	1.57
Phoenix Of Spain IRE	20	117	17	1.50
Cracksman GB	53	397	13	1.45
Sioux Nation USA	96	753	13	1.44
Calyx GB	20	124	16	1.42
Too Darn Hot GB	28	191	15	1.40
Saxon Warrior JPN	60	499	12	1.39
Study Of Man IRE	8	51	16	1.27
Zoustar AUS	68	637	11	1.21
Seahenge USA	7	49	14	1.21
Wings Of Eagles FR	6	37	16	1.19
Rajasinghe IRE	20	169	12	1.17
Expert Eye GB	69	671	10	1.14
Ten Sovereigns IRE	19	174	11	1.14
Soldier's Call GB	29	268	11	1.12

Figure 3: indices of stallions with two racing crops in 2023

Sire	Damsire	Horses	Wins	Runs	xW	p value
Galileo IRE	Danehill USA	182	254	1445	236	0.213
Galileo IRE	Danehill Dancer IRE	93	126	705	108	0.067
Dutch Art GB	Pivotal GB	75	125	906	118	0.459
Fastnet Rock AUS	Galileo IRE	72	106	567	77	0.001
Dubawi IRE	Galileo IRE	77	98	460	92	0.485
Dark Angel IRE	Oasis Dream GB	58	95	729	88	0.393
Galileo IRE	Darshaan GB	68	92	540	90	0.817
Oasis Dream GB	Sadler's Wells USA	53	86	640	79	0.400
Oasis Dream GB	Selkirk USA	49	80	537	69	0.137
Zoffany IRE	Galileo IRE	69	80	599	65	0.056
Invincible Spirit IRE	Pivotal GB	40	75	467	56	0.008
Fastnet Rock AUS	Sadler's Wells USA	57	74	529	68	0.476
Invincible Spirit IRE	Galileo IRE	47	74	427	55	0.009

Figure 4: successful sire-damsire combinations judged by total wins since 2010

Are nicks real and which are the most prominent?

Adding together the independent effect of sire and broodmare could easily underestimate certain combinations of sire and damsire known as nicks.

To test for nicks, we can extract every stallion–damsire combination from the data, predict the observed strike rate in the data from their independent effects, then test whether any difference is statistically significant – how likely is it to have happened by chance? This probability is known as a p-value.

Since 2010, there have been 49,724 different sire-damsire combinations expressed by runners on the Flat in Britain and Ireland. Figure 4 shows the most highly successful combinations, together with the p-value - the probability that the difference between our prediction for wins (**xW**) and actual wins (**Wins**) is down to chance.

For example, using Odds Ratios as we did in Figure 1, runners by Dutch Art out of a Pivotal mare should have won 118 races. This cohort did win 125 races – 7 more than expected – but over a sample of 906 races this is 48.5% likely to happen just at random. So, we must be at least a little sceptical that there is anything special about the combination considering that both Dutch Art and the deceased Pivotal are good sires anyway. It is only natural that their combination should produce good results!

By contrast, the nicks represented by the sire-damsire crosses of Galileo-Danehill Dancer, Fastnet Rock-Galileo, Invincible Spirit-Pivotal and Invincible Spirit-Galileo far exceeded expectation and are most unlikely the result of mere chance. In these cases, we can conclude that a positive interaction takes place when sire and damsire combine.

Finally, Figure 5 lists the 20 most unusually successful sire-damsire combinations, judged by their associated p-values.

Sire	Damsire	Horses	Wins	Runs	xW	p value
New Approach IRE	Street Cry IRE	11	25	86	12	0.000
Fastnet Rock AUS	Galileo IRE	72	106	567	77	0.001
Dubawi IRE	Dansili GB	29	58	203	39	0.001
Galileo IRE	Pivotal GB	31	73	314	50	0.001
Sea The Stars IRE	Kingmambo USA	16	38	120	23	0.001
Dubawi IRE	Singspiel IRE	28	50	170	33	0.001
Kingman GB	Selkirk USA	12	26	98	14	0.002
Deep Impact JPN	Galileo IRE	13	22	85	12	0.002
Dark Angel IRE	Lawman FR	12	22	96	11	0.002
Exceed And Excel AUS	Elusive Quality USA	27	50	273	33	0.003
Dandy Man IRE	Marju IRE	13	34	208	20	0.003
Dubawi IRE	Shamardal USA	29	47	173	32	0.004
Frankel GB	Dubawi IRE	30	58	207	41	0.005
Fast Company IRE	Pivotal GB	12	30	182	18	0.006
Teofilo IRE	Dubawi IRE	26	41	184	27	0.006
Invincible Spirit IRE	Pivotal GB	40	75	467	56	0.008
Distorted Humor USA	A.P. Indy USA	12	24	97	14	0.009

Figure 5: successful sire-damsire combinations most likely to constitute a genuine nick

Conclusion and further thoughts

This is a fascinating topic. Of the 49,724 crosses in the database, only 3.5% meet the criterion of statistical significance at the 5% level. So, this exercise is some proof that nicks do exist but are a lot rarer than many bloodstock people think.

Working with this data has convinced me there is a logical thread to pedigrees which has not yet been widely explained. I think that great stallions arise mostly because they have an unusual capacity to form successful nicks with female families which are themselves potent. Through the stud book is then this cascading effect which results in some lines being strengthened and others dying out.

In other words, rather than mere genetic potency, a stallion's ability to form a successful outcross from previously dominant lines results in not just lots of opportunities for combination with talented mares to which other stallions are too closely related.

There is a genetic interaction which results in the probability of a foal reaching a super-high threshold like 115 greatly increases. Meanwhile, for other crosses between elite sires and elite broodmares, the effect is merely additive and the potential for a high-class foal increases linearly – not exponentially.

This is evident in the record of Galileo and Danzig-line mares, for instance. In turn, finding a stallion who interacts exponentially with mares by Galileo has been prominent in the ambitions of the Coolmore brains-trust for some time. Could that sire be Justify?