Skill and Chance in Association Football

By C. Reep and B. Benjamin

In association football the ball is passed from player to player among the eleven members of a side until a particular player loses possession of the ball either by interception or tackle on the part of a member of the defending team or by an infringement of the rules of the game or by himself shooting at the defending side's goal. We may define an "r-pass movement" as one in which a player of side A, having just obtained the ball, sets off a series of r successful passes among members of his own team after which there is either a shot at goal by the rth recipient or an infringement or there is an attempted $(r+1)$th pass which is intercepted. We note that a 0-pass movement means that A's first attempted pass is intercepted or that there is a shot at goal without a preceding pass, as, for example, from the penalty spot. There are a number of factors affecting the likelihood of a successful rth pass:

(1) the positions of the players between whom the pass is attempted and the defending players who try to intercept;

(2) the relative skills of the players and the effectiveness and confidence with which those skills are applied at this particular stage of the game.

In evenly matched teams playing under the conditions normally obtaining in good class football (for example, in the first three English F.A. Divisions) the second of these factors does not vary widely from one attempted pass to another but as the attack proceeds the opponents progressively dispose themselves to improve the chance of interception or tackle; they are in more compact formation and closer to the ball. The probability of an r-pass movement $P(r) = [p_1 . p_2 . \ldots . p_j . \ldots . p_r(1-p_{r+1})]$ where $p_1 > p_2 > p_3 \ldots p_r > p_{r+1}$.

As to the form of the function p_r one would expect p_1 to be fairly high though less than unity and p_r to fall rapidly to some low value beyond which there is little further decrease; an exponential form seems likely.

One of us (C. R.), who has an acute interest in the implications for strategic training, has compiled careful records of actual frequencies against which to test this theory. Table 1 shows four observed distributions together with values of p_r derived from these distributions. It will be seen that generally p_r declines as r increases, if somewhat irregularly. At higher values of r when numbers are smaller the observed ratios $\{P(r-1)\}/\{P(r-2)\}$ become irregular and this produces even greater irregularity in p_r.

Investigation of the data suggested that $P(r)$ would be a more convenient parameter to attempt to "smooth" in order to produce a standard model. In contemplating this "smoothing" and the shape of the distribution $P(r)$, it seemed from inspection that one would expect $P(r)$ to fit closely to a negative binomial distribution and this expectation was justified.

Table 2 shows the result of attempting to fit negative binomials to the distributions of Table 1. Generally these are reasonably good fits; the fit for the miscellaneous

TABLE 1

Frequency of r-pass moves in F.A. matches and derived values of $p_r = 1 - [[\{P(r-1)\}/\{P(r-2)\}]\{(1/p_{r-1}) - 1\}]$

r	42 *First Division matches 1957–58*†		12 *First Division matches 1961–62*‡		36 *Miscellaneous matches 1965–66*		11 *World Cup matches 1966*§	
	No. of moves	p_r	*No. of moves*	p_r	*No. of moves*	p_r	*No. of moves*	p_r
0	10,580	·445	2,331	·380	3,407	·387	1,862	·368
1	6,923	·555	1,629	·620	2,320	·613	1,220	·632
2	3,611	·475	1.035	·571	1,435	·570	785	·619
3	1,592	·424	565	·523	767	·534	518	·604
4	608	·400	304	·502	422	·533	316	·567
5	280	·427	154	·466	215	·518	153	·534
6	107	·381	60	·420	130	·526	78	·577
7	33	·380	31	·463	62	·456	65	·626
8	9	·497	15	·400	24	·431	27	·503
9+	11		6		20		39	
Total	23,754		6,130		8,802		5,063	
Mean	1.00		1.30		1·35		1·56	
Variance	1·50		2·11		2·52		3·29	

† All involving Sheffield Wednesday.
‡ All involving Arsenal.
§ One match went to extra time.

TABLE 2

Frequency of r-pass moves in F.A. matches—negative binomial tests

r	42 *First Division matches 1957–58*†			12 *First Division matches 1961–62*‡			36 *Miscellaneous matches 1965–66*			11 *World Cup matches 1966*§		
	Actual	*Expected*	$A-E$	*Actual*	*Expected*	$A-E$	*Actual*	*Expected*	$A-E$	*Actual*	*Expected*	$A-E$
0	10,580	10,542	+38	2,331	2,224	+107	3,407	3,341	+66	1,862	1,777	+85
1	6,923	7,075	−152	1,629	1,788	−159	2,320	2,407	−87	1,220	1,313	−93
2	3,611	3,542	+69	1,035	1,061	−26	1,435	1,428	+7	785	829	−44
3	1,592	1,572	+20	565	556	+9	767	786	−19	518	503	+15
4	608	653	−45	304	271	+33	422	415	+7	316	293	+23
5	280	260	+20	154	126	+28	215	215	0	153	165	−12
6	107	101	+6	60	57	+3	130	109	+21	78	93	−15
7	33	38	−5	31	25	+6	62	55	+7	65	52	+13
8	9	14	−5	15	11	+4	24	27	−3	27	29	−2
9 and over	11	8	+3	6	10	−4	20	22	−2	39	31	+8
Total	23,754	23,805		6,130	6,129		8,802	8,805		5,063	5,085	
Mean	1·00			1·303			1·348			1·558		
Variance	1·495			2·109			2·519			3·286		
$P(\chi^2)$			>·10			<·0001			>·30			<·01

† Involving Sheffield Wednesday on each occasion.
‡ Involving Arsenal on each occasion.
§ One match went to extra time.

series is very close while that for the Arsenal series is bad. But these are only four examples chosen arbitrarily from the extensive records shown in Table 3.

TABLE 3

Passing move distributions 1953–67

Detail	*No. of matches*	*Frequency*							
		0	1	2	3	4	5	6	7+
1953–54 Wolverhampton W.	12	·417	·299	·169	·071	·029	·009	·004	·002
1955–56 Miscellaneous	15	·395	·295	·167	·081	·036	·017	·005	·004
1955–56 Sheffield W.	42	·433	·297	·158	·068	·027	·011	·004	·002
1956–57 Sheffield W.	42	·443	·297	·152	·064	·027	·010	·005	·002
1956–57 Miscellaneous	18	·446	·296	·145	·062	·029	·013	·006	·003
1957–58 Sheffield W.†	42	·445	·291	·152	·067	·025	·013	·005	·002
1957–58 Miscellaneous	12	·414	·284	·158	·076	·036	·018	·007	·007
1958 World Cup	11	·384	·262	·161	·090	·051	·027	·012	·013
1958–59 Miscellaneous	33	·398	·290	·170	·077	·036	·017	·008	·004
1959–60 Miscellaneous	41	·395	·296	·159	·080	·039	·017	·007	·007
1960–61 Miscellaneous	38	·389	·266	·160	·090	·049	·025	·012	·009
1960–61 Tottenham H.	12	·395	·259	·156	·092	·049	·028	·012	·009
1961–62 Tottenham H.	12	·378	·260	·157	·094	·055	·031	·012	·013
1961–62 Arsenal†	12	·380	·266	·169	·092	·050	·025	·010	·008
1961–62 Miscellaneous	42	·378	·262	·166	·092	·051	·028	·012	·011
1961–62 Burnley	8	·384	·275	·175	·083	·040	·022	·012	·009
1962–63 Tottenham H.	11	·400	·251	·160	·092	·050	·026	·012	·009
1962–63 Miscellaneous	42	·396	·268	·160	·086	·045	·024	·012	·009
1962–63 Coventry	14	·434	·294	·151	·067	·032	·014	·005	·003
1962 World Cup	18	·330	·245	·170	·110	·066	·036	·020	·023
1963–64 Miscellaneous	19	·391	·268	·153	·091	·047	·027	·013	·010
1964–65 Miscellaneous	17	·366	·263	·167	·093	·056	·025	·014	·016
1965–66 Miscellaneous†	36	·387	·263	·163	·087	·048	·025	·015	·012
1966 World Cup†	11	·368	·241	·155	·103	·062	·030	·015	·026
1966–67 Miscellaneous	18	·369	·231	·163	·102	·057	·033	·023	·022
All matches—Actual	578	·394	·275	·162	·084	·044	·022	·010	·009
Expected‡		·398	·275	·158	·084	·043	·022	·011	·009
A − *E*		−·004	—	+·004	—	+·001	—	−·001	—

† Shown in Tables 1 and 2.
‡ On the hypothesis of a negative binomial.

In this table the frequency distributions have been standardized to a total frequency of unity. Their general similarity is striking. Taking all matches together, the fit to the negative binomial distribution is almost exact, and this is the pattern adhered to by most of the series from 1958 to 1967. Prior to 1958, the negative binomial is still representative though the parameters are those of the first distribution shown in Table 2. General time changes in the style of play would be expected to affect the basic parameters of the distribution but not its mathematical character.

There are other regularities to be observed in association football, and some of these are shown in Table 4. If the pitch is divided into four quarters laterally then the

TABLE 4

Season	*Total matches*	*Detail*	*Ratio of shooting area origin goals to all goals*	*Average No. of shots to score 1 goal*	*Ratio of shots from regained possessions to all such moves*	*Ratio of shots from shooting area origin attacks to all such attacks*	*Ratio of all shots to all attacks reaching the shooting area*	*Ratio of goals "against" conceded by "own half" breakdowns to all goals against*	*Ratio of shooting area origin attacks to all attacks reaching the shooting area*	*Ratio of regained possessions in shooting area to all shooting area origin attacks*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1952–53	26	Miscellaneous	·624	10·1				·516		
1953–54	51	Miscellaneous	·528	9·4				·503		
1954	8	World Cup	·308	9·1	·314	·231	·170	·359	·245	48·0
1954–55	43	Miscellaneous	·494	9·0				·466		
1954–55	14	Sheffield W.	·510	9·8	·296	·223	·125	·449	·301	52·5
1955–56	15	Miscellaneous	·397	9·8	·319	·250	·157	·349	·296	54·3
1955–56	42	Sheffield W.	·558	9·6	·327	·244	·137	·515	·287	52·2
1956–57	18	Miscellaneous	·533	10·8	·266	·211	·133	·567	·329	52·8
1956–57	42	Sheffield W.	·535	9·5	·299	·230	·140	·500	·299	51·6
1957–58	42	Sheffield W.	·547	9·0	·282	·218	·128	·472	·313	52·3
1957–58	15	Miscellaneous	·546	8·0	·293	·228	·153	·468	·299	54·7
1958	11	World Cup	·525	9·7	·297	·225	·141	·525	·313	53·1
1958–59	36	Miscellaneous	·538	7·9	·267	·206	·133	·506	·292	51·3
1959–60	41	Miscellaneous	·544	9·4	·272	·217	·138	·513	·310	52·0
1960–61	38	Miscellaneous	·556	8·9	·291	·222	·149	·469	·305	52·8
1961–62	42	Miscellaneous	·546	10·6	·265	·208	·141	·511	·299	52·7
1962	18	World Cup	·520					·340		
1962–63	42	Miscellaneous	·551	10·4	·275	·216	·143	·497	·299	51·8
1963–64	26	Miscellaneous	·536	8·8	·275	·216	·146	·409	·299	53·6
1964–65	17	Miscellaneous	·480	11·8	·257	·213	·146	·540	·265	48·6
1965–66	50	Miscellaneous	·575	10·0						
1966	11	World Cup	·600	16·2	·246	·218	·152	·560	·311	50·9
1966–67	18	Miscellaneous	·426	10·4	·269	·221	·153	·475	·293	52·1

Note: Blank cells in this table indicate that the necessary records were not compiled for the particular set of matches.

two quarters containing the goals are referred to as the "shooting areas". As might be expected more goals come from passing moves beginning in that area; the proportion can be seen from column (4) to be consistently just about 50 per cent. Column (10) shows that such shooting area origin attacks are in the ratio of 1 : 3 to attacks reaching the shooting area, and from column (11) we see that of these shooting area origin attacks slightly more than 50 per cent are regained possessions, that is, the defence fails to get the ball clearly away from the shooting area. The size of these proportions are not surprising but their near constancy is remarkable. Again column (6) shows that 30 per cent of regained possessions lead to shots at goal.

It can be seen from column (8) that 15 per cent of all attacks reaching the shooting area lead to shots at goal, but from column (7) that of attacks originating in the shooting area 22 per cent or so lead to shots at goal.

Column (9) refers to "own half" breakdown, that is, to passing moves which break down in the attackers' own half of the pitch. Of all goals scored against them 50 per cent come from such failures to move the ball into the defenders' half.

Finally, column (5) shows that with rare exceptions (for example, the 1966 World Cup series) it takes 10 shots to score 1 goal.

The observation that there is a stochastic element in the number of goals arising from a particular number of shots in one match (as well as a near-constant proportion over a larger series of matches) is easy for a statistician to accept; indeed he would be surprised if it were otherwise. It indicates, of course, that an excess of shots by one team does not mean that, by chance, the other side will not get more goals and thus win the match. All this is so far removed from current soccer beliefs and tactics that general acceptance of the random element has been inhibited (though one of us, C. R., has shown that a successful style of play can be built upon it). It seems, however, that chance does dominate the game and probably most similar ball games.