
AGENDA 9.3

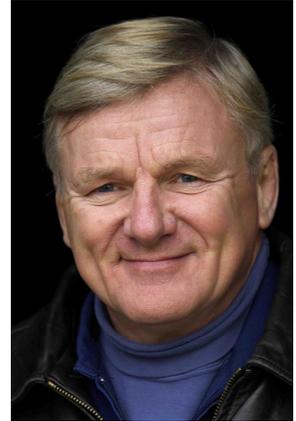
CIVA CATALOGUE SUB-COMMITTEE REPORT

Alan Cassidy, Chairman

**The Difficulty Coefficients of Rolling Turns in the Aresti System
(Condensed)**

“The rolling circle is the absolute king in aerobatics.” - Eric Müllerⁱ

The sad state of this monarchy, however, is that the king is paid a pauper’s wage.



Executive Summary

This report exposes inconsistencies and illogical grading of K factors resulting from the current method for evaluation of difficulty coefficients of Rolling Turns. The current K factors are unfair. Pilots who fly difficult Rolling Turns well are not rewarded for their efforts.

Areas of doubt expressed by CIVA Delegates over the consistency of judging figures from Family 2 (Rolling turns) are investigated by analysis of data from WAC 2005. These fears are shown to be ungrounded.

A revised methodology for evaluating the coefficients for Rolling Turns is more fully described. Its adoption by CIVA is recommended.

The Purpose of Coefficients

The purpose of aerobatic competition is to find the best pilot. Pilots are sometimes told which figures to fly. At other times, they choose the figures for themselves. Some figures are more difficult to fly and some figures are easier to fly. All pilots would find that difficult figures need more skill to fly than easy figures. It is essential that the Aresti System defines the relative difficulty of figures in a logical and consistent way.

The difficulty coefficients (K) describe the relative difficulty of performing particular figures., The difficulty coefficients are NOT a measure of how easy or difficult figures are to judge and judging issues should NOT affect the assignment of difficulty coefficients to each figure

Difficult Figures will generally be graded lower than easier figures:

Across a group of pilots, more difficult figures will be graded lower, on average, than easier figures, because pilots are more likely to make mistakes when they fly more difficult figures. However, if a pilot flies a difficult figure very well, (s)he should get a higher grade than a pilot

ⁱ Müller, Eric 1983. **Flight Unlimited**, Müller & Carson, ISBN 0 9509252 0 9, p129.

that flies the same figure badly. All pilots must expect, on average, to be graded lower for difficult figures than for easy ones.

Difficult Figures must have higher k factors than easier figures

If one pilot flies a difficult figure and gets a grade of 8 marks and another pilot flies an easier figure and also gets a grade of 8 marks, the first pilot should get more points once the grades have been multiplied by the difficulty coefficients (K factor) for each figure.

Difficult Figures are more discriminating than easier figures

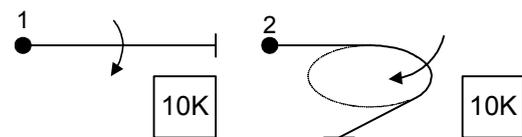
In any contest, it is the more difficult figures that will show differences between the more skilled and less skilled pilots.

Key Point: The pilot who flies a difficult figure better than his competitors is entitled to expect the appropriate numerical reward.

The Current Situation

The current method of determination for the difficulty coefficients of Rolling Turns (Aresti Family 2.1 – 2.20) is described in Part II, Paragraph 7 of the Aresti catalogueⁱⁱ. We consider the method used inappropriate, as it leads to a number of situations which are clearly absurd.

It is the view of the CIVA Catalogue Sub-Committee that to perform an aileron roll on a turning flight path is more difficult than to perform the same roll in a straight line. This opinion is not contradicted by any pilot that we know. However, the current method of evaluation for rolling turns leads to the following examples.

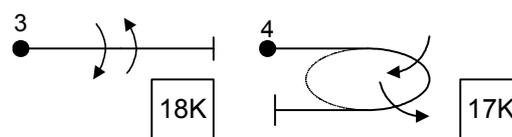


Example A

In Example A, Figure 2 is more difficult than Figure 1 yet they have the same overall difficulty coefficient (K Factor). Thus, if a pilot flies the harder figure more accurately, he does not get the appropriate reward:

	Figure 1	Figure 2	Score
K	10	10	
Pilot 1	8.0	7.0	150
Pilot 2	7.0	8.0	150

The pilots have the same overall score, even though Pilot 2 has flown the harder figure with more skill. ***The better pilot has not received the appropriate reward.***

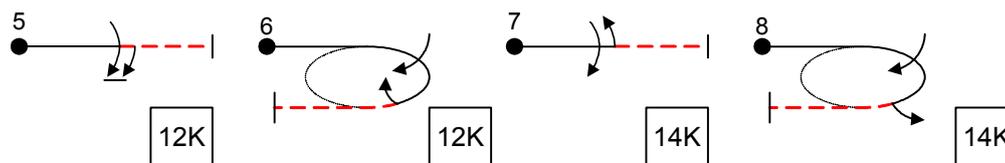


Example B

In Example B, Figure 4 is much harder to fly than Figure 3 and would usually be graded lower. But it is possible for the better pilot to actually lose this competition like this:

	Figure 1	Figure 2	Score
K	18	17	
Pilot 1	8.0	4.0	212
Pilot 2	7.0	5.0	211

There are many more examples...



Example C

Key Point: These examples clearly demonstrate that the current system of evaluation of rolling turns is inappropriate.

Judging Issues

The Catalogue Sub-Committee's view is that the K Factor for any manoeuvre should truly reflect the difficulty of flying the figure and be totally independent of any issue about the quality of judging.

Some delegates to CIVA expressed a number of concerns, at the 2006 Plenary, about the way that judges assessed rolling turns. These concerns can be summarised as:

1. Judges find rolling turns more difficult to judge than other figures.
2. Grades (figure marks) awarded for rolling turns are much less consistent than grades given to other figures.

And therefore:

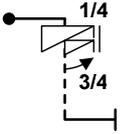
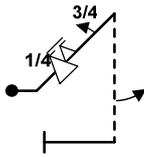
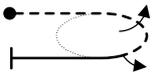
3. If judges give inconsistent grades to rolling turns, the K factors should not be increased because the figure grades given are less reliable than the grades given to other figures.

Most CIVA delegates who voted against changing the K factors for Rolling turns did so because they had concerns about the way rolling turns are judged, not because they disagreed with the logic of the revisions proposed by the Judging Sub-Committee.

These concerns will be addressed here. The data from WAC 2005 was available in a spreadsheet ready for analysis, so this data was used for ease of preparing this report. We have used the judging of three figures from the Programme Q at the WAC 2005 for this analysis. The Q sequence is a good one to use for a number of reasons:

ⁱⁱ Aresti, Colonel José Luis. **Aresti Aerobatic Catalogue (Condensed)**, Aresti System SL, 1961-2005, ISBN 84 4040996 6

- The sequence has been well practiced by all pilots and so performance should be a true reflection of pilots' relative abilities.
- There were 48 pilots so that there is a good prospect of a statistically robust analysis of what happened.
- There were ten judges, so the analysis of the agreement between the judges is more robust than if fewer judges were involved. This analysis looks at three figures from the WAC Q Programme:

Figure			
Figure No.	3 - Spin	4 - Wedge	6 – Rolling Turn
Coefficient	20K	40K	13K

The Rolling Turn (Figure 6) was chosen because this report concerns Rolling Turns.

The Spin (Figure 3) was chosen for comparison because it, too, has relatively complex judging criteria (stalled entry, constant rotation, exit attitude etc).

The Wedge (Figure 4) was chosen because it is 40K which is typical of the coefficient that would be attributed under the new system to a 4-roll, 360° rolling turn and also has complex judging criteria.

We are now going to look at the judging of these three figures and ask two questions:

- How does the pilot performance vary, from best pilot to worst pilot, for each figure?
- How consistent are the judges at agreeing whether a figure is good or bad?

Variation in Pilot Performance

Pilot performance varies and judge performance also varies. The field of WAC pilots is quite large and there are significant differences in skill levels between the best pilots and the least experienced. It is likely that this difference in skill level will be discovered more noticeably in difficult figures requiring particular skill. Hence the variability of each judge's grades for these figures will demonstrate how much difference the judges see between well-flown and below-average figures.

One of the particular strengths of the Fair Play System, is that it analyses all the raw grades for a particular figure in a single data set. In the Q Programme the Fair Play System analyses the results for each figure as a separate group. Analysing the results in Figure Groups allows the system to make allowance for the different judging styles for the different types of figures. So, for example, the analysis lets us look at how the panel of judges as a whole assessed just the rolling turns of all pilots. Similarly, the grades for the Spin (Figure 3) or the Wedge (Figure 4) can also be analysed separately. The first stage of the FPS process is to normalise the raw

grades so that each judge has the same average grade and the same spread (deviation) of grades. During the normalisation quoted here, Soft Zeroes are kept at zero. This is how FPS will work from 2007 onwards. The grouping of grades for each single figure produces a data set that looks like this:

Figure f – Normalised Data

	Judge 1	Judge 2	Judge 3	Judge 4	...	Judge j		
Pilot 1	Grade	Grade	Grade	Grade	...	Grade	Pilot Average	Deviation for Pilot 1
Pilot 2	Grade	Grade	Grade	Grade	...	Grade	Pilot Average	Deviation for Pilot 2
Pilot 3	Grade	Grade	Grade	Grade	...	Grade	Pilot Average	Deviation for Pilot 3
....		
....		
....		
Pilot p	Grade	Grade	Grade	Grade		Grade		
	Judge Average	Judge Average	etc	etc	etc	etc	Group Average	Average of Pilot Deviations
	Judge Deviation	Judge Deviation	etc	etc	etc	etc	Group Deviation	

How well do pilots fly/score each figure?

The **Group Average**, in the yellow box, is the average normalised grade for the Figure across all pilots and all judges. It shows the average figure grade on a scale of 0 (worst) to 10 (best). In the WAC 2005 Q Programme, the Group Averages were as follows:

		Group Average (Average Figure Grade across all pilots)
Figure 3	Spin (K=20)	7.16
Figure 4	Wedge (K= 40)	6.43
Figure 6	Roller (K=13)	6.16
Average over all Figures		6.77

Pilots were given the highest grade for the spin and the lowest grade for the rolling circle. This could reflect real differences in the quality of figures flown, differences in how judges assess the figures, or a combination of both factors. The average figure grades suggest that the spin was flown best (and, by inference, an easier figure to fly) and the rolling circle was flown worst (a more difficult figure to fly). Most pilots would agree that the spin is easier to fly than the rolling circle. Most pilots would also agree that both the Wedge and the Rolling Circle are more difficult to fly than the spin. Both of these figures are given lower average grades than the spin, which suggests that judges are seeing pilots make more mistakes in the difficult figures and are

deducting marks to reflect these errors. The average grades suggest that the figures are listed in increasing level of difficulty:

Easiest: Spin
Wedge
Most Difficult: Rolling Circle

Yet the rolling turn currently carries the **lowest** coefficient of the three.

Do judges assess the Spin, Wedge and Rolling Turn in different ways?

The Group Deviation, in the orange box, shows the average Standard Deviation of the normalised grades. This shows the overall spread in grades given to each figure across all judges and all pilots. A high Standard Deviation indicates that the grades given to a particular figure were very varied (wide range of grades and more differences in opinion amongst the judges). In the WAC 2005 Q Programme, the Group Deviations were as follows:

Group Deviation (Average variability in grades for each figure)		
Figure 3	Spin	1.32
Figure 4	Wedge	1.73
Figure 6	Roller	1.41
Average over all Figures		1.40

These data show that the judges' grades were most variable for the Wedge (Figure 4) and least variable for the Spin (Figure 3). Grades for Rolling turns were, on average, much less variable than grades for the Wedge. The variations in grades for spins and rolling turns were very similar. Judges saw more difference between the best and worst examples of Figure 4 than between the best and worst rolling turns.

The Deviation for each Pilot, in the lilac boxes, shows the level of disagreement between the judges about the quality of each pilot's figure. If judges were finding it particularly hard to apply the grading criteria, as some think for the rolling turns, then this figure would be noticeably high.

The Average of Pilot Deviations, in the red box, shows the average of the pilot deviations and gives a measure, as a whole, of how consistent the panel of judges are at evaluating each individual figure. Again, the larger this number, in comparison with the same statistic for other figures, the less consistent are the judges for this figure. In the WAC 2005 Q Programme, The Average of Pilot Deviations were as follows:

Figure 3	Spin	0.94
Figure 4	Wedge	1.15
Figure 6	Roller	1.06
Average over all Figures		0.95

The judges are rather more consistent in judging the spin than the rolling turn, but they are much less consistent at judging the wedge. This data contradicts the suggestion that judges are in poor agreement over the quality of rolling turns.

Key Point: There was no particular inconsistency in judging Rolling Turns at WAC. In fact, the Rolling Turn was more consistently judged than a typical 40K figure.

Data Used

Appendix 3, at the end of this report, shows the normalised data from which the numbers above are extracted. The normalisation has been done in accordance with the slightly revised process which will be published in CIVA Regulations, Part 1, 2007. In this type of normalisation, soft zero grades are excluded from the normalisation process and remain unchanged up to the point where anomalies are assessed and replaced.

Summary of Judging Analysis

This analysis leads to the following summary of conclusions:

- Rolling Turns are confirmed as being difficult, as the average normalised grade is relatively low.
- Rolling Turns show a greater difference in pilot skill level, between the best and the worst, than simpler figures. They are therefore more discriminating and should play a greater part in determining the selection of the best pilot.
- Judges are a little less consistent at judging Rolling Turns than simple spins, but they are more consistent with rolling turns than a typical higher K figure (the Wedge).

The analysis has found NO evidence that judges grade Rolling turns significantly differently to other figures that are equally difficult to fly.

There is NO evidence that judges are especially bad at assessing the quality of Rolling turns.

Concerns about how judges assess rolling turns are not a valid reason for rejecting the proposed revision to the difficulty coefficients.

Key Point: Judging differences, that might mitigate against the proper application of suitable K-factors for Rolling Turns, are shown NOT to exist in the data group analysed.

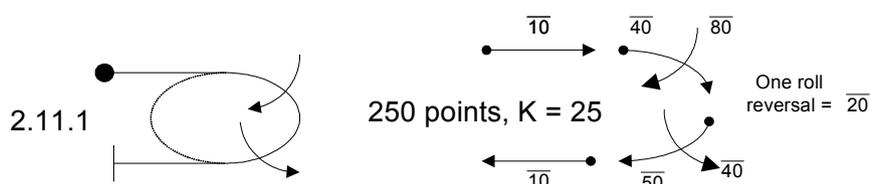
A Logical and Consistent Method of Determining the Difficulty Coefficients for Rolling Turns

We have shown earlier that the current method of evaluation of coefficients for Rolling Turns was inappropriate by comparison with similar straight-line rolling figures. It is therefore necessary to propose a revised methodology that is not only internally logical but also consistent with other methodologies in the Catalogue.

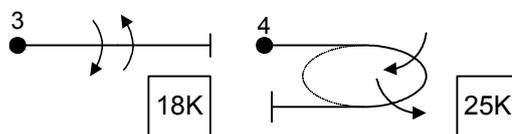
The proposed methodology is as follows:

- The horizontal entry and exit lines but contribute an element as for every other figure in the Catalogue. This is 10 points for an upright line and 13 points for an inverted line
- The methodology must contain an element for each 90° arc of turn. This is the first major difference between straight and turning figures and must be recognised in the method. As turning is achieved with continually variable inputs of rudder and elevator, it is much more difficult than in a simple steep turn. When rolling outwards the difficulty of the turn is greater than when rolling inwards, primarily because, as the rudder is changed from one side to the other, the newly applied rudder tends to pull the nose down rather than holding it up.
 - For inward rolls, each 90° arc of turn is accorded 40 points (50 in glider aircraft)
 - For outward rolls, each 90° arc of turn is accorded 50 points (70 in glider aircraft)
- There must be an element for the first roll in the figure. To be consistent with the evaluation of Family 9.1, this roll is allocated 80 points (120 in gliders) which is the same as 9.1.3.4. This coefficient applies to the first roll in a 90° arc of turn.
- When a roll is placed on a 120° arc of turn, it must be performed with slower roll rate and the difficulty of turning, especially during the wings-level phases, increases. Therefore a roll on a larger arc of turn is more difficult. The degree of difficulty increases in a non-linear manner as the extent of the arc increases. Therefore the first roll on these larger arcs will attract a higher number of points as follows:
 - For the first roll on an arc of 120°, 106 points (150 in gliders)
 - For the first roll on an arc of 180°, 160 points (240 in gliders)
 - For the first roll on an arc of 360°, 280 points (420 in gliders)
- Any second and subsequent rolls count 50% of the first roll, as is done in the case of straight linked rolls. In the exceptional instance of Sub-Family 2.16, the half roll is evaluated as a full roll.
- When alternate rolls occur, stops are required as in hesitation rolls and then the direction of roll must be reversed. For each reversal of direction a further 20 points (50 in gliders) are added.
- As everywhere else in the catalogue, the total sum of points for the figure is divided by 10 and then rounded up or down to give the final coefficient.

The method of evaluation is exemplified in the figure below.



Thus in the proposed modified Catalogue, the following comparison would be drawn:



Example D

Key Point: The proposed revision produces much more logical and appropriate coefficients than the current situation (see Example B).

The Effect of Increased K for Rolling Turns

The following beneficial effects will be seen from the allocation of appropriate, increased coefficients for Rolling Turns, details of which are shown at Appendices 1 and 2.:

- **A pilot flying a truly difficult figure better than his competitors will receive the just reward for his higher level of skill.**
- **In Free Programmes, where a figure from Family 2 is required, the aggregate K needed for the non-Family 2 figures will reduce. This will go some way to countering the “Category Creep” feared by some Delegates as a result of reducing the number of Advanced Free figures from 15 to 12.**
- **The intellectual integrity of our sport will be enhanced by eliminating absurdities such as those illustrated at the beginning of this paper.**

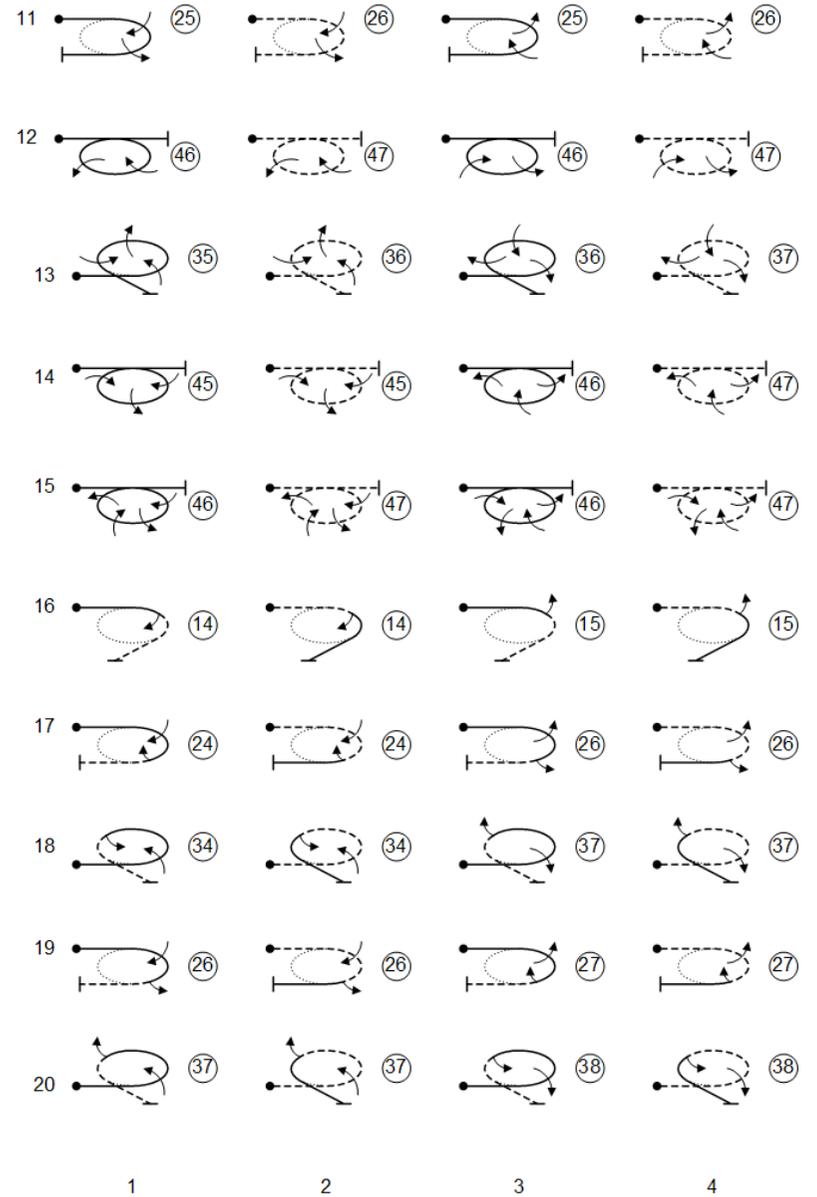
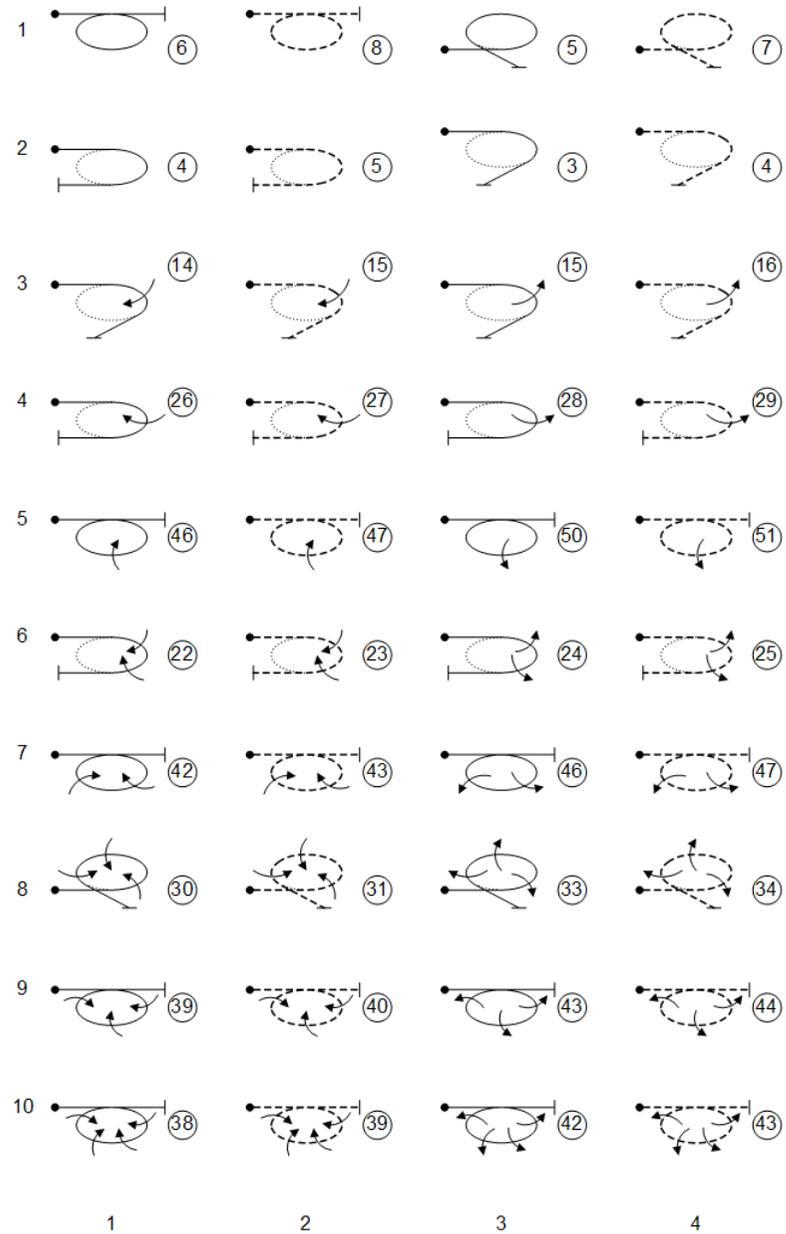
The Judging inconsistencies, feared by some Delegates and also by some pilots, have been shown here to be insignificant.

Recommendation

All Delegates of CIVA are urged to study this report carefully and to discuss its implications with the pilots in their Associations. You are then urged to vote in favour of the proposed changes at the 2007 CIVA Plenary Meeting.

Alan Cassidy
UK Delegate
Chairman, Catalogue Sub-Committee

Appendix 1 - Revised K-factors for Power Family 2.



Appendix 2 - Revised K-factors for Gliders Family 2.

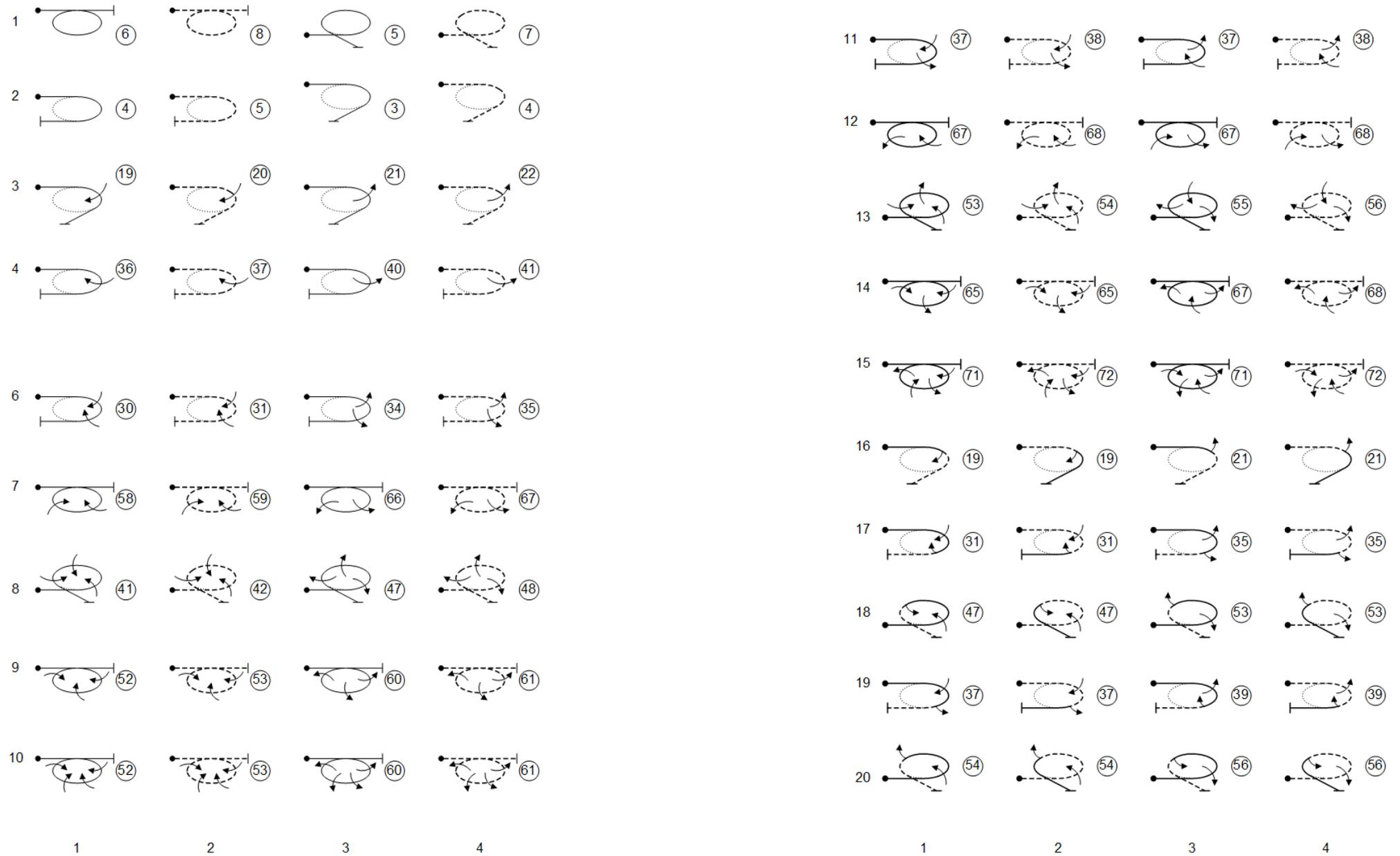


Figure 4 - Wedge

J1	J2	J3	J4	J5	J6	J7	J8	J9	J10	
7.85	7.95	8.20	6.62	0.00	6.95	8.98	6.92	7.91	6.91	2.51
6.25	6.35	7.67	8.11	7.21	7.41	5.55	4.90	7.91	6.91	1.05
6.79	6.75	7.15	7.62	7.21	7.41	6.53	6.92	6.54	6.34	0.41
7.85	6.35	8.73	8.11	7.94	7.86	8.49	7.59	7.23	6.91	0.72
5.72	6.75	7.15	6.62	7.94	7.41	6.04	6.92	4.50	5.78	1.00
7.32	7.15	6.62	7.62	7.21	7.41	8.00	6.92	7.23	5.78	0.60
7.32	7.55	6.10	7.12	6.85	6.95	6.53	7.59	5.86	7.48	0.61
7.85	7.95	8.20	8.11	8.30	7.86	8.00	6.92	7.23	8.05	0.44
6.79	7.55	7.15	7.12	6.49	6.04	7.02	5.57	6.54	7.48	0.63
6.25	0.00	7.15	5.63	7.21	5.59	6.04	3.55	6.54	6.91	2.20
0.00	0.00	M	M	5.03	1.49	0.00	M	M	M	2.18
8.38	7.15	7.15	8.11	6.49	7.41	6.53	6.25	7.23	8.05	0.73
8.38	7.55	7.15	7.62	7.58	6.95	7.02	6.25	7.23	7.48	0.56
4.66	5.95	5.57	6.62	6.85	6.50	6.04	5.57	6.54	6.34	0.66
6.25	4.76	5.04	7.62	7.21	6.95	7.02	2.87	5.18	6.34	1.46
5.19	M	4.52	5.63	5.39	6.04	4.08	6.25	6.54	6.34	0.85
6.79	5.95	7.15	4.64	4.67	6.95	6.04	5.57	7.23	6.91	0.98
6.25	4.76	7.15	0.00	7.21	6.50	7.02	6.92	5.86	6.34	2.17
6.79	7.15	5.04	6.62	6.12	6.95	8.00	7.59	7.23	8.05	0.90
8.38	4.76	6.10	6.13	3.57	6.95	6.04	7.59	5.18	0.00	2.37
7.32	7.55	7.67	7.62	7.94	7.41	8.00	8.27	7.23	8.05	0.35
8.38	7.55	8.73	8.61	8.30	7.86	8.00	6.92	7.91	8.05	0.53
7.32	0.00	8.20	6.13	6.49	7.86	7.51	8.27	7.91	M	2.59
8.38	4.76	7.15	6.13	7.58	7.41	8.98	8.27	7.91	5.78	1.31
5.72	6.75	6.10	5.13	5.76	7.41	6.04	7.59	6.54	7.48	0.84
6.25	6.75	7.15	6.13	3.57	6.95	6.04	7.59	6.54	3.50	1.41
7.32	7.55	7.15	7.62	6.85	7.41	7.02	8.94	6.54	6.34	0.72
7.85	7.15	7.15	6.62	6.85	6.95	8.00	8.94	7.23	7.48	0.68
6.79	7.55	7.67	7.62	7.21	6.95	7.02	6.92	7.23	8.05	0.41
6.25	3.16	7.15	7.62	7.21	6.95	6.04	7.59	7.23	6.91	1.32
5.19	5.55	2.94	4.14	0.00	4.22	4.08	6.25	3.14	6.34	1.88
5.19	6.75	7.15	7.12	6.85	6.50	5.55	6.92	6.54	6.91	0.66
6.79	0.00	5.04	0.00	4.67	6.04	6.04	6.92	7.23	2.37	2.76
6.25	M	6.10	6.62	4.30	6.50	6.04	6.25	5.86	5.78	0.68
4.66	4.76	5.57	4.14	7.94	6.04	6.04	5.57	7.23	5.78	1.15
6.25	7.15	5.04	5.63	5.39	6.50	6.04	6.92	6.54	6.91	0.70
7.32	7.15	7.15	7.12	7.94	7.86	7.02	6.92	7.91	8.05	0.44
2.52	3.96	3.99	3.65	0.00	4.22	4.08	5.57	5.86	3.50	1.63
7.32	7.15	5.04	7.62	6.85	6.95	6.53	4.90	7.23	6.91	0.93
7.32	7.55	7.15	7.62	7.58	6.50	7.51	7.59	7.23	7.48	0.34
0.00	7.95	7.15	5.13	6.49	6.50	6.04	6.25	0.00	6.91	2.86
5.19	7.95	6.62	6.62	0.00	7.41	7.02	5.57	6.54	6.91	2.25
6.25	7.55	7.15	6.13	6.49	6.95	6.04	6.25	6.54	5.78	0.55
7.32	7.15	7.67	7.62	7.21	6.50	8.00	7.59	7.23	6.91	0.43
7.85	7.95	7.15	7.62	8.30	7.41	7.02	6.25	7.91	8.05	0.61
M	M	M	M	M	M	M	M	M	M	
5.19	0.00	6.10	4.64	5.76	3.31	4.08	0.00	1.77	6.34	2.39
7.32	7.55	7.15	7.62	7.21	7.86	8.49	6.92	7.91	6.91	0.50
6.40	5.94	6.68	6.39	6.11	6.68	6.54	6.53	6.53	6.53	6.43
1.81	2.41	1.22	1.82	2.22	1.22	1.55	1.56	1.56	1.56	1.73

