The Valet Score:

Principles

November 2015
Søren Hein
soren.hein@gmail.com

Table of Contents

1 Motivation

1.1 SCORING

Ted Turner, the U.S. media mogul, reportedly once said, "*Life is a game. Money is how we keep score.*"

Bridge is maybe not life, but in bridge we also keep score. Pairs tournaments are often scored at matchpoints. Team tournaments are often scored at IMPs converted into VP (victory points).

In a sense, the official scoring is all that matters. And yet some bridge players are analytically inclined and want to know more, whether to identify weaknesses in their own game or to have the satisfaction of seeing that it was all their teammates' fault…

There are various ways in which one can go about getting more detail. Surely one of the best ways is to conduct post-mortems with strong players who take a balanced view of things. But there are also analytical methods that can at least give pointers, the Butler score for team games being perhaps the best-known.

Any method, including the official scoring method of a given tournament, will have limitations, and it is important to be clear about these. If you bid a good slam that needs one of two finesses and you go down, you might consider yourself moderately unlucky. If you lose the tournament because of this – well, you can cry into your beer afterwards, but you still lost. At least at this level we accept that statistics and "luck" play a role.

Perhaps it is helpful to think of analytical methods as follows.

Results	▪ Official score ▪ Butler ▪ Valet	(not common, e.g. Epson tournament)
Cards	▪ Double dummy ▪ Simulations	▪ Post-mortem ▪ Single dummy
	Quantitative	**Manual**

Quantitative methods are not subject to judgment, whereas manual analysis attempts to find the "truth." Maybe it didn't cost that you missed that trump lead, but you should still have found it. So in the post-mortem your friends should charge you for this, and you should learn from it – maybe next time it *will* matter.

Result-based methods look only at the actual scores achieved at the table, without considering the actual distribution of the cards. Maybe you bid a game on a finesse for the king of trumps and went down, and others stayed out of it. Or maybe you bid the same game missing instead the ace of trumps, so the contract never had any play. You cannot tell the difference between these two distributions just by looking at the scores. For that you need card-based analysis which goes beyond the actual scores.

In this document I aim to introduce and explain the Valet score, an improvement and generalization of the Butler score. I want to be clear about what you can and cannot expect from the Valet score.

You *can* expect

- Quantitative scores for bidding, play and defense (including optionally a separate score for the opening leads). Play and leads are scored for each player separately, while the rest is scored for the pair.
- Statistical averages. If you bid to a good contract that fails, you will get a poor official score and also a poor Butler and Valet score. This is supposed to average out over a sufficiently large number of hands.
- Scores that are based on the actual results of real players.

You *cannot* expect

- Anything that relates directly to the actual distribution of the cards, except if this is visible in the results.
- Double-dummy results.
- Perfection or truth.

1.2 HISTORY

I don't actually know when and by whom the Butler score was introduced. I'd like to think that it was a Mr. or Mrs. Butler. But if you're seriously interested in scoring (and you might not be reading this document otherwise), you should perhaps be aware that it is a bit sneaky to speak of *the* Butler score. As typically implemented, it involves a datum score which may or may not exclude outliers, and which may be an average or a median value or something else. Different scoring programs implement Butler scoring differently; for example, the current WBF program excludes the top and bottom two scores in the Bermuda Bowl. To get a flavor for the considerations, check out

http://www.bridgeguys.com/pdf/butler_scoring_stevenson.pdf

My own interest was piqued in 2004 after I played in the Istanbul Olympiad for Luxembourg, with more valor than success. I realized that the tournament offered a high-quality source of relatively homogeneous data, with many comparisons per board, and over the Christmas vacation I came up with the Valet score principles.

I then got the hand results from pdf documents at ECats, and for reasons that probably made sense to me at the time, I implemented the Valet score in VBA macros within Microsoft Excel. I wrote a short article about it and submitted it to the Bridge World. The article was not accepted – I suspect that Jeff Rubens does not share my view of the potential value of quantitative methods outside of the official scoring. After that I forgot about the Valet score.

On October 14, 2015, the issue came up on Bridge Winners:

http://bridgewinners.com/article/view/recalculating-butler-scores-to-delineate-declaring-vs-defending-proficiency/

I took the opportunity to post my description from 2004:

http://bridgewinners.com/article/view/the-valet-score/

and I was encouraged by the interest, public as well as private. So I took the leap and rewrote and generalized the code in a more practical way. With this document I am releasing an open-source project in C++. It consists of a stand-alone program and a DLL / library that can be called from other programs.

This document explains the principles. A companion documents explains the computer program. I hope that authors of scoring programs will incorporate the Valet score, just as they do the Butler score today.

Oh, the name? "Valet" means "jack" in French, and a valet has some similarity with a butler.

I would like to thank the people who have commented on the concept and on the documents, particularly Thorvald Aagaard who offered many helpful comments and who was also the first user of the program.

2 Definition

2.1 THE BUTLER SCORE

It is easiest to explain the Butler score is with an example. Thorvald Aagaard brought up the below hand from the 2015 Bermuda Bowl in Chennai. Nothing fancy – on the North-South hands you want to be in slam. The actual 22 results were as shown below.

East dealer NS vul.	♠ A 4 ♡ J 7 6 5 4 2 ♢ A 8 ♣ A Q 4	
♠ J 9 8 3 2 ♡ Q T ♢ J 7 2 ♣ 6 5 2	N W E S	♠ K T ♡ 9 3 ♢ 9 5 4 3 ♣ K T 8 7 3
	♠ Q 7 6 5 ♡ A K 8 ♢ K Q T 6 ♣ J 9	

http://www.worldbridge.org/repository/tourn/chennai.15/Microsite/Asp/BoardAcross.asp?qboard=002.01..1130

Contract	Count	Results (lead)
4NT, S	2	1 × 11 (♠), 1 × 12 (♣)
4♡, S	8	8 × 12 (4 × ♠, 4 × ♣)
4♡, N	3	3 × 12 (2 × ♢ , 1 × ♣)
6♡, S	7	7 × 12 (3 × ♠, 4 × ♣)
6♡, N	2	12 (♢) , 13 (♣)

The first step is to calculate the datum score. In the simplest version we just take the average of the scores, so we have 1 × 1460 + 8 × 1430 + 1 × 690 + 11 × 680 + 1 × 660 which comes out to 21730. Dividing by 22 yields 988. The second step is to convert each score into a Butler score.

Score	Difference NS	Butler score NS
1460	472	+10 IMPs
1430	442	+10 IMPs
690	-298	-7 IMPs
680	-308	-7 IMPs
660	-328	-8 IMPs

By the way, you can reproduce this result with the valet program by using the `example1.txt` and `names1.txt` files in the distribution and invoking the program as "`valet.exe –n names1.txt –s example1.txt`" (setting the directory as well with the `–d` option if necessary).

There are a number of variations on the datum calculation.

- We may leave out, for example, the highest and lowest score from the average calculation on the view that the extreme scores should not "distort" the "real" average. In the present case this only changes the datum score into 981, but the effect can of course be larger as well. My view is that bridge is about all of these scores, and at least in a strong field these large deviations are part of the overall outcome. So I am not in favor of this adjustment personally.
- We may round the datum score either to the closest multiple of 10, or always round it down. In the example, we could round 988 to 980 (down) or to 990 (nearest). It turns out not to matter here, but the IMP scale has gaps – what is the IMP score for a score difference of 894 against the datum?
- We may not calculate the arithmetic average, but the median value, i.e. the value "in the middle" if we order the values. As we have an even number of results, we might calculate the average of the two middle scores. In the example, both these scores are 680. This does make a big difference to the IMP scores, of course. Note that the sum of the IMP scores in the table above is -2 for NS and hence +2 for EW. This is only about +0.1 IMPs on average, but it is not zero. If the datum is 680, the average NS IMP score against the median datum is +5.3 IMPs. My view is that the sum should be close to zero (ideally zero), and I am not in favor of the median datum.

What should we think about the Butler score in this case? East-West are largely innocent bystanders – they don't have a suitable preempt, and they will just get the opposite score from the North-South pair that they happen to catch. Is this "fair"? No, but it is supposed to average out over time.

It is also possible to calculate a Butler-like score using IMPs across the field (which I sometimes abbreviate to "IAF"). It is a matter of debate whether this is then still a Butler score or not. It is more work, but it is not a problem for computers.

Let us start with the declarer who scored +1460 for bidding slam and making the overtrick. S would score +30 against the 8 declarers who made +1430 (1 IMP), +770 (13 IMPs) against the single declarer making +690, +780 against 11 declarers with +680 (also 13 IMPs), and +800 against a solitary +660 (also 13 IMPs). In total this comes to +177 IMPs in 21 comparisons, or +8.43 IMPs on average. The full set of IAF scores is shown below.

Score	**Butler score NS**
1460	+8.43 IMPs
1430	+7.95 IMPs
690	-5.14 IMPs
680	-5.52 IMPs
660	-6.14 IMPs

Note that the sum of all 22 results is now exactly 0 IMPs. Also note that the IAF values are lower than the Butler datum IMPs. This is a general phenomenon that we will come back to. The datum score first calculates a linear average and then applies the nonlinear IMP scale (a difference of 400 does not result in twice the IMP score of a 200 difference). The IAF method first applies the nonlinear IMP scale to all differences and then calculates a linear average. The IMP scale is compressive, so it can be shown mathematically that the "IAF IMPs" yield lower numbers than the "datum IMPs". It turns out that the sum of the IAF scores for the two pairs in a team is closer to the team performance than is the case if we add up the datum-IMP Butler scores.

You can reproduce the above results with the valet program by using the `example1.txt` and `names1.txt` files in the distribution and invoking the program as "`valet.exe –v imps –n names1.txt –s example1.txt`".

2.2 THE VALET SCORE

The Valet concept applies to Butler datum scoring, but also to other scoring forms. I will first explain the concept using datum scoring.

How do we separate the bidding from the play/defense, looking only at the results? The key is to imagine that you bid to your actual contract, but then you are called away and replaced by an average declarer of your chosen denomination (notrump, spades etc.) who plays on your behalf. In fact you are replaced by all the declarers playing in your denomination from your side (let's call it notrump for example), one after the other, and you get the average of their results. It turns out to be mathematically convenient to include yourself as one of these declarers, as some things that ought to come out to zero on average really do behave in this way then.

If all these notrump declarers get better results than the datum score, then you would think that you've bid to a superior contract. The difference between this average result and the datum score is your bidding score, and whatever is left between your bidding score and your Butler score is your play score.

The defenders get the opposite of your bidding score as their bidding score, and the opposite of your play score as their defense score. This may not be fair in a particular case, but it is no more unfair than the actual score or the Butler score.

In order to get enough data, we take all the notrump declarers, even the ones who played at a different level than you. If you bid game, we still include those declarers who played in one notrump. You might say that the one-notrump declarer could play differently than the three-notrump declarer, and you would be right. But we do need enough data points, and many hands are played the same way no matter what, including our slam hand above.

We only consider the declarers who played in your denomination from your side. If you sat North and protected your king, while the South players were vulnerable to a lead through that king, then rest assured that you will normally get a good bidding score.

Let us imagine that you bid to 3NT on some hand, and three declarers bid 3NT, two of them (perhaps including yourself) making 9 and 1 making 8 tricks. There were also three declarers bidding 1NT, two of them making 9 and 1 making 8 tricks. We would then give you the score for being in 3NT making 9 tricks 3 times and making 8 tricks 3 times in 3NT. Even though one declarer made 8 tricks in 1NT, we still apply his result to your 3NT contract, and you are penalized for his play – tough luck.

The above enables us to split the Butler score into two components: Bidding and play/defense. Can we go even further and grade the opening lead separately from the rest of the defense? You will be unsurprised to hear that the answer is a qualified "yes". Maybe in a later version we'll distinguish between spot cards and honors, at least.

We are now starting to slice the data rather finely, and we are not going to consider individual cards of a suit separately. It may be that you brilliantly led an unsupported queen, squashing declarer's stiff jack. The pleasure will have to be its own reward.

We are going to look at all the pairs who were on lead against a contract in your denomination from your side (notrump). We already know the average score that all these opening leaders achieved against notrump, as this is the opposite of the bidding score that we assigned to you above. We are now going to use the same principle and to calculate the same type of score, but only against the opening leaders who led the suit that you were faced with (call it clubs).

Once we have this score, we are going to compare it to the total bidding score. If the average club leader did better than his bidding score, then this is considered to be due to his club lead. Whatever is left of his defense score is then his subsequent-defense score.

2.3 SCORING FORMS

If we are using datum scoring, then each comparison is made by first calculating an average score over the universe of opponents that we are considering. For example, if we let all the notrump declarers play for you, then we take their average score as the datum. We then compare each of the scores that these declarers would have achieved for you, convert it into IMPs against this notrump datum, and in the end we average over the declarers.

It is also possible to use IAF scoring. In this case we compare each of the declarers to each of the opponents separately, convert each score into IMPs, and then average these.

Personally I think that IAF scoring is the only valid choice. With datum scoring it often happens that an allocation into bidding and play produces a small, but non-zero value for a component that really "should" be zero. It's a bit more work for the computer and for the programmer, but so be it.

It is also possible to use matchpoint scoring! This means that we can also decompose a matchpoint tournament into bidding, play, lead and defense components. Matchpoints are slightly different in that they are not additive. At IMPs it is reasonable to think of the

Butler or Valet score as the sum of a number of components. At matchpoints it is an average.

For example, say that you get a 60% score on your board. You actually bid to a contract that the average declarer would have played for 70%, so your bidding score is 70%. We would like the average of your bidding and play score to be 60%, so this must mean that your play was only worth 50%. In general your play score is twice your overall IAF score, minus your bidding score (here: twice 60% minus 70% equals 50%). As Thorvald points out, it is easier to think of MPs as IMPs that are constained to +1, 0 and -1.

There is actually no particular reason that your matchpoint score should be the average of the two. Implicitly we are saying that the two disciplines are equally important. But it as an intuitively pleasing choice.

In extreme cases it is possible for the play score on a single hand to be less than 0% or more than 100%! This will even out over enough hands, though.

2.4 EXAMPLES

Let's go back to our slam hand and look at the declarer who scored +1430 for bidding and making the heart slam. We are only going to consider the 15 declarers who played in hearts from South. Every last one of them made exactly 12 tricks. So your average score over all declarers (including yourself) is also +1430. Some of the others played in game, but you still get +1430 for their 12 tricks as we apply their tricks to your contract.

Recall that your +1430 was worth +10 datum IMPs (and +7.95 IAF IMPs) overall.

We are now going to score this synthetic score against the datum. Of course this yields the same result as your actual +1430 score, i.e. +10 datum IMPs. So your bidding score is +10 and your play score is 0 IMPs. This makes sense. You didn't outplay anybody, but you did bid to the right contract.

It turns out to be the same for the ones staying in 4 hearts – zero play score.

Let us look at the single declarer who played in 4NT making only 11 tricks. There were 2 notrump declarers, one making 11 and one making 12. The scores against the datum were -8 and -7 IMPs, respectively, so the average score was -7.5 IMPs. This is the bidding score for both declarers. The one who made 12 tricks gets +0.5 IMPs and the one who made only 11 tricks gets -0.5 IMPs for declarer play.

You might say that the overwhelming number of declarers made 12 overall, but most of them were not in notrump, and we can't consider that. It is also clear that you want to be in hearts in this particular case.

In IAF terms it comes out slightly differently, of course. The +1460 pair gets +8.43 IMPs consisting of + 8.05 bidding IMPs and +0.38 play IMPs. In general it comes out as shown below. In my view this is maybe not perfect, but it is pretty close.

Score	Butler score NS	Bid / play
1460	+8.43 IMPs	+8.05 / +0.38
1430	+7.95 IMPs	+7.95 / 0.00 (South) +8.05 / -0.10 (North)
690	-5.14 IMPs	-5.64 / +0.50
680	-5.52 IMPs	-5.52 / 0.00 (South) -5.28 / -0.24 (North)
660	-6.14 IMPs	-5.65 / -0.50

This was a hand that was mostly about the bidding. Let's do a couple more. The first one (Example 2) is from the Polish first league in 2013/14, Round 1, Segment 1, Board 1 and is concerned with the play. The hand was played 16 times, in every single case in 3NT from North, so that makes it easier. The results were as shown below.

http://www.pzbs.pl/wyniki/liga/liga2013-14/ekstraklasa/e1314rr1b-1.html

North dealer
None vul.

North:
♠ K 6 2
♡ A K
♢ Q J 8 5 3
♣ A J 5

West:
♠ 9 5 4
♡ Q T 9 6
♢ K
♣ K 8 7 6 3

N
W E
S

East:
♠ A Q T
♡ J 4 3 2
♢ T 9 7 4
♣ 9 4

South:
♠ J 8 7 3
♡ 8 7 5
♢ A 6 2
♣ Q T 2

Lead	Count	Results
♠	2	1 × 11, 1 × 9
♡	6	4 × 9, 2 × 7
♢	8	6 × 9, 1 × 8, 1 × 7

The datum score was (12 × 400 + 1 × (-50) + 3 × (-100)) / 16 = 278. Therefore the IMPs were as follows for North-South.

Score	Datum	IAF
+460	+5	+4.40
+400	+3	+2.73
-50	-8	-7.67
-100	-9	-8.93

The Valet score allocates the IMPs as follows, for the datum and IAF IMPs separately. In all cases the scores are given from the players' point of view, positive being good.

Score	Lead	Overall	Bid	Play
+460	♠	+5	+0.19	+4.81
+400	♠	+3	+0.19	+2.81
	♡	+3	+0.19	+2.81
	♢	+3	+0.19	+2.81
-50	♢	-8	+0.19	-8.19
-100	♡	-9	+0.19	-9.19
	♢	-9	+0.19	-9.19

Datum scoring, declarer's view

Score	Lead	Overall	Bid	Lead	Other def.
+460	♠	-5	-0.19	-3.81	-1.00
+400	♠	-3	-0.19	-3.81	+1.00
	♡	-3	-0.19	+1.19	-4.00
	♢	-3	-0.19	+0.06	-2.88
-50	♢	+8	-0.19	+0.06	+8.13
-100	♡	+9	-0.19	+1.19	+8.00
	♢	+9	-0.19	+0.06	+9.13

Datum scoring, defenders' view

Score	Lead	Overall	Bid	Play
+460	♠	+4.40	0.00	4.40
+400	♠	+2.73	0.00	+2.73
	♡	+2.73	0.00	+2.73
	♢	+2.73	0.00	+2.73
-50	♢	-7.67	0.00	-7.67
-100	♡	-8.93	0.00	-8.93
	♢	-8.93	0.00	8.93

IAF scoring, declarer's view

Score	Lead	Overall	Bid	Lead	Other def.
+460	♠	-4.40	0.00	-3.57	-0.83
+400	♠	-2.73	0.00	-3.57	+0.83
	♡	-2.73	0.00	+1.16	-3.89
	♢	-2.73	0.00	+0.03	-2.76
-50	♢	+7.67	0.00	+0.03	+7.64
-100	♡	+8.93	0.00	+1.16	+7.78
	♢	+8.93	0.00	+0.03	+8.91

IAF scoring, defenders' view

This is the first time that we see meaningful results from play, lead and defense, so we should reflect on these results. Let's start with datum scoring.

From declarer's point of view, most of the scores are allocated to the play. This makes sense. But why is there a +0.19 datum-IMP score for bidding 3NT when everybody else also bids it?

Let's take the lone declarer who made 11 tricks. We are going to replace him with each of the 16 declarers, score their results against the datum score, convert to IMPs and then add up the scores. In this way we get $1 \times 5 + 11 \times 3 + 1 \times (-8) + 3 \times (-9) = +3$ IMPs. The average bidding score is therefore 3/16 = 0.19 IMPs. You will recall that the same thing happened in Example 1 above – in general, the total number of IMPs for one side does not have to add up to 0 with datum scoring.

This problem does not occur with IAF IMP scoring, as every combination of declarer and defenders occurs exactly twice during IMPs-across-the-field: Once from the point of view of the declarer, and once from the defenders' side. Indeed we see that IAF scoring assigns 0.00 IMPs to all bids in this particular case. This is one reason that I consider IAF scoring to be superior, and I'm not going to try to justify datum scores from here on.

We also observe again that IAF IMPs are lower than datum IMPs.

Let's look at the scores associated with different leads. When a spade was led, one declarer made 11 and another made 9 tricks. We already know that the one whose declarer made 11 tricks got -4.40 IMPs for his trouble, while the other spade leader got away with -2.73 IMPs. The Valet score allocates the average value, -3.57 IMPs, to their lead and the difference to their other defense. That seems reasonable.

Let's look at the red-suit leaders. There were six heart leaders, and four paid out 9 tricks while two scored down two. So the defenders' IAF IMPs are $4 \times (-2.73) + 2 \times 8.93 = 6.94$ IMPs or an average of 1.16 IMPs.

Of the eight diamond leaders, six suffered 9 tricks while one obtained down one and one scored down two. The defenders' IAF IMPs are $6 \times (-2.73) + 1 \times 7.67 + 1 \times 8.93 = -0,22$ IMPs or an average of -0.03 IMPs. The diamond leaders beat the contract twice out of eight times, and one of those was only down one. The heart leaders beat the contract twice out of only six times, both times down two. It makes sense that the heart leaders get the better lead score.

For practice, the actual output table from the Valet program for this one hand is shown on the next page. The first six columns are hand numbers: Total, number of declarer players for players 1 and 2, number of defenses and number of opening leaders for players 1 and 2. Then comes the overall score. Then come the component scores.

In this example there is only one hand. In general the average are shown *per hand played*, not per hand declared, led or defended. Otherwise the scores would not easily add up.

Players	No.	Pl1	Pl2	Def	L1	L2	IMPs	Bid	Play1	Play2	Lead1	Lead2	Def
C Serek - P Busse	1	0	0	1	0	1	8.93	-0.00	0.00	0.00	0.00	0.03	8.91
R Woliński - M Urbański	1	0	0	1	0	1	8.93	-0.00	0.00	0.00	0.00	1.16	7.78
R Kierznowski - L Sztyrak	1	0	0	1	1	0	8.93	-0.00	0.00	0.00	1.16	0.00	7.78
Z Papierniak - J Figlus	1	0	0	1	1	0	7.67	-0.00	0.00	0.00	0.03	0.00	7.64
A Suwik - M Kowalski	1	1	0	0	0	0	4.40	0.00	4.40	0.00	0.00	0.00	0.00
R Jagniewski - W Gaweł	1	1	0	0	0	0	2.73	0.00	2.73	0.00	0.00	0.00	0.00
R Nowicki - R Stoliński	1	0	1	0	0	0	2.73	0.00	0.00	2.73	0.00	0.00	0.00
W Starkowski - P Gawryś	1	1	0	0	0	0	2.73	0.00	2.73	0.00	0.00	0.00	0.00
D Filipowicz - K Martens	1	1	0	0	0	0	2.73	0.00	2.73	0.00	0.00	0.00	0.00
J Ciechomski - M Leśniewski	1	1	0	0	0	0	2.73	0.00	2.73	0.00	0.00	0.00	0.00
A Jeleniewski - J Wachnowski	1	0	1	0	0	0	2.73	0.00	0.00	2.73	0.00	0.00	0.00
K Jassem - M Mazurkiewicz	1	1	0	0	0	0	2.73	0.00	2.73	0.00	0.00	0.00	0.00
K Pikus - M Cichocki	1	0	1	0	0	0	2.73	0.00	0.00	2.73	0.00	0.00	0.00
L Ohrysko - M Jeleniewski	1	0	1	0	0	0	2.73	0.00	0.00	2.73	0.00	0.00	0.00
A Guła - M Taczewski	1	0	1	0	0	0	2.73	0.00	0.00	2.73	0.00	0.00	0.00
G Bajek - M Wręczycki	1	0	1	0	0	0	2.73	0.00	0.00	2.73	0.00	0.00	0.00
P Żak - J Zaremba	1	0	0	1	0	1	-2.73	-0.00	0.00	0.00	0.00	0.03	-2.76
S Gołębiowski - A Żmudziński	1	0	0	1	1	0	-2.73	-0.00	0.00	0.00	1.16	0.00	-3.89
M Dembiński - M Pędziński	1	0	0	1	0	1	-2.73	-0.00	0.00	0.00	0.00	0.03	-2.76
G Narkiewicz - K Buras	1	0	0	1	0	1	-2.73	-0.00	0.00	0.00	0.00	0.03	-2.76
J Kalita - M Nowosadzki	1	0	0	1	0	1	-2.73	-0.00	0.00	0.00	0.00	1.16	-3.89
T Wiśniewski - I Kowalczyk	1	0	0	1	0	1	-2.73	-0.00	0.00	0.00	0.00	-3.57	0.83
G Głasek - M Piwowarczyk	1	0	0	1	1	0	-2.73	-0.00	0.00	0.00	1.16	0.00	-3.89
T Jochymski - M Kania	1	0	0	1	1	0	-2.73	-0.00	0.00	0.00	0.03	0.00	-2.76
C Komajda - J Kotorowicz	1	0	0	1	0	1	-2.73	-0.00	0.00	0.00	0.00	1.16	-3.89
M Bartkowski - M Puczyński	1	0	0	1	1	0	-2.73	-0.00	0.00	0.00	0.03	0.00	-2.76
R Matlak - Z Guła	1	0	0	1	0	1	-2.73	-0.00	0.00	0.00	0.00	0.03	-2.76
J Klukowski - A Kowalski	1	0	0	1	1	0	-4.40	-0.00	0.00	0.00	-3.57	0.00	-0.83
S Zawiślak - E Miszewska	1	1	0	0	0	0	-7.67	0.00	-7.67	0.00	0.00	0.00	0.00
B Pazur - T Pilch	1	0	1	0	0	0	-8.93	0.00	0.00	-8.93	0.00	0.00	0.00
L Niemiec - M Białożyt	1	1	0	0	0	0	-8.93	0.00	-8.93	0.00	0.00	0.00	0.00
P Nawrocki - P Wiankowski	1	1	0	0	0	0	-8.93	0.00	-8.93	0.00	0.00	0.00	0.00

Actual Valet output for Example 2

West dealer
All vul.

	♠ T 3 ♡ A Q ♢ K Q 5 ♣ A K Q J 7 4	
♠ Q 9 7 6 4 2 ♡ 5 3 ♢ 8 7 6 ♣ 6 2	N W E S	♠ 8 ♡ J 7 6 ♢ J T 9 2 ♣ T 9 8 5 3
	♠ A K J 5 ♡ K T 9 8 4 2 ♢ A 4 3 ♣ -	

The next hand (Example 3) is from the Polish first league in 2013/14, Round 1, Segment 1, Board 4. The hand was played 16 times and the results are shown in the table below.

http://www.pzbs.pl/wyniki/liga/liga2013-14/ekstraklasa/e1314rr1b-4.html

Contract	Declarer	Count	Lead	Tricks
7NT	S	4	2 × ♠, 1 × ♡, 1 × ♣	13
7NT	N	3	1 × ♠, 2 × ♢, 1 × ♣	13
6NT	N	1	1 × ♢	13
5NT	N	1	1 × ♢	13
7♡	S	3	1 × ♡, 2 × ♢	13
6♡	N	2	2 × ♠	12, 13
7♣	N	1	1 × ♡	12
6♣	N	1	1 × ♣	12

We would expect there to be nothing to the play. Indeed we get the IAF-IMP tables shown on the next page. The two pairs who played in clubs both made 12 tricks. One single heart declarer made only 12 tricks, but in play terms this didn’t cost very much, as he was only in 6 and there were 10 pairs in grand slam and one pair only in game, where the missing trick made no difference in IMPs.

Score	Lead	Overall	Bid	Play
+2220	♠, ♡, ♣	+5.93	+5.93	0.00
+2210	♡, ♢	+5.80	+5.80	0.00
+1470	♢	-6.20	-6.20	0.00
+1460	♠	-6.47	-6.60	+0.13
+1430	♠	-6.73	-6.60	-0.13
+1370	♣	-7.33	-7.33	0.00
+720	♢	-13.53	-13.53	0.00
-100	♡	-18.67	-18.67	0.00

IAF scoring, declarer's view in Example 3

Score	Lead	Overall	Bid	Lead	Other def.
+2220	♠, ♡, ♣	-5.93	5.93	0.00	0.00
+2210	♡, ♢	-5.80	-5.80	0.00	0.00
+1470	♢	+6.20	+6.20	0.00	0.00
+1460	♠	+6.47	+6.60	0.00	-0.13
+1430	♠	+6.73	6.60	0.00	+0.13
+1370	♣	+7.33	+7.33	0.00	0.00
+720	♢	+13.53	+13.53	0.00	0.00
-100	♡	+18.67	+18.67	0.00	0.00

IAF scoring, defenders' view in Example 3

North dealer NS vul.	♠ K J 9 2 ♡ K J T 7 5 3 ♢ 9 3 ♣ 3	
♠ T 4 3 ♡ 6 ♢ Q 7 5 ♣ K Q 9 7 6 4	N W E S	♠ A 8 7 6 ♡ 9 8 4 2 ♢ T ♣ A J 8 5
	♠ Q 5 ♡ A Q ♢ A K J 8 6 4 2 ♣ T 2	

The next hand (Example 4) is from the Polish first league in 2013/14, Round 1, Segment 1, Board 5. The results are all over the place, including a sacrifice, so here we see what the Valet score does about that.

http://www.pzbs.pl/wyniki/liga/liga2013-14/ekstraklasa/e1314rr1b-5.html

Contract	Declarer	Count	Lead	Tricks
6♢	S	1	1 × ♣	10
5♢	N	1	1 × ♣	10
5♢	S	4	1 × ♣	10
3♢	S	1	1 × ♡	9
5♣×	E	1	1 × ♢	9
4♡	N	7	2 × ♠, 2 × ♡, 3 × ♣	12 (1 × ♣) 11 (other)
4♡	S	1	1 × ♣	12

The overall IAF IMPs are as follows.

Score	IAF
+680	+6.93
+650	+6.33
+300	+0.33
+110	-4.07
-100	-7.67
-200	-9.13

The Valet tables are shown on the next page. The format is different as the lead is now more relevant.

Contract	Lead	Tricks	Count	Overall	Bid	Play
4♡, N	♠	11	2	+6.33	+6.42	-0.09
	♡	11	2	+6.33	+6.42	-0.09
	♣	12	1	+6.93	+6.42	+0.51
		11	2	+6.33	+6.42	-0.09
4♡, S	♣	12	1	+6.93	+6.93	0.00
5♣×, E	♢	9	1	+0.33	+0.33	0.00
6♢, S	♣	10	1	-9.13	-9.29	+0.15
5♢, S	♣	10	4	-7.67	-7.91	+0.24
3♢, S	♡	9	1	-4.07	-3.49	-0.58
5♢, N	♣	10	1	-7.67	-7.67	0.00

IAF scoring, declarer's view in Example 4

Contract	Lead	Tricks	Count	Overall	Bid	Lead	Other def.
4♡, N	♠	11	2	-6.33	-6.42	+0.09	0.00
	♡	11	2	-6.33	-6.42	+0.09	0.00
	♣	12	1	-6.93	-6.42	-0.11	-0.40
		11	2	-6.33	-6.42	-0.11	+0.20
4♡, S	♣	12	1	-6.93	-6.93	0.00	0.00
5♣×, E	♢	9	1	-0.33	-0.33	0.00	0.00
6♢, S	♣	10	1	+9.13	+9.29	-0.15	0.00
5♢, S	♣	10	4	+7.67	+7.91	-0.24	0.00
3♢, S	♡	9	1	+4.07	+3.49	+0.58	0.00
5♢, N	♣	10	1	+7.67	+7.67	0.00	0.00

IAF scoring, defenders' view in Example 4

The play differences are small. All declarers but one made 10 tricks in diamonds, so the odd one out got -0.58 IMPs for his play (or for the opponents' heart lead). The total IAF IMPs for this result are -3.49 IMPs and the play score is the balance of -0.49 IMPs.

Hold on, you say – why -3.49 IMPs? That is not in the table of IAF IMPs for actual scores on the previous page. That's true. But we have to imagine the other five declarers in diamonds from South. They all made 10 tricks, effectively scoring +130 for our declarer, and this is a score we haven't seen yet. We then have to calculate the IAF IMPs for +130 vs. the field.

Which field? When we do a normal IAF calculation, we drop the player whom we're scoring. But here the +130 result comes from one player's contract and another player's declarer play. It turns out that we get more intuitively pleasing results in some cases (and it's easier in my computer implementation) if we drop one player with the score that we’re using. In this case there is nobody with a +130 score, so we drop nobody, and we compare with the full field of 16 players. There is no right or wrong here, I think.

So this comes to 2 × (-11) + 6 × (-11) + 1 × (-5) + 1 × 1 + 5 × 6 + 1 × 8 = -54 IMPs for the scores +680, +650, +300, +20, -100, -200, and the average of these 16 results is -3.38 IMPs. So we get 1 × (-4.07) + 5 × (-3.38) = +20.94 IMPs divided by 6 declarers, which comes to -3.49 IMPs.

Well, that was exciting – you can see why we need a computer for this. Looking at the defenders, their only chance to make a difference after the lead was in the case of a club lead against four hearts by North. The remaining variation was due to the lead, which makes sense.

North dealer
None vul.

North:
♠ K J T 6
♡ K 8 7 5
♢ 8 3 2
♣ Q 2

West:
♠ 9 8 5
♡ A Q 9 3
♢ J T 7
♣ 9 8 3

N
W E
S

East:
♠ A 4 3
♡ J T 4 2
♢ A K Q 9
♣ A J

South:
♠ Q 7 2
♡ 6
♢ 6 5 4
♣ K T 7 6 5 4

http://www.worldbridge.org/repository/tourn/chennai.15/Microsite/Asp/BoardAcross.asp?qboard=001.01..1130

As our penultimate example, let's take another hand (Example 5) brought up by Thorvald Aagaard from the 2015 Bermuda Bowl (Round 1, Board 1). Thorvald says that East-West should be in four hearts as it mainly depends on the heart finesse. Ideally he would like the score to reflect this, whether or not the king is in place. In any event he would like most of the variation to be in the play and defense, not in the bidding.

As discussed in the introduction, the Valet score cannot take into account a specific king, as it has no way to incorporate actual distributions, only the results from actual play. But it's a hand where the defense can slip up. Against 3NT the defense has to lead a club. Against four hearts the defenders have to avoid some pitfalls, including a club ruff in East. So let's see what happens (next page).

First of all, 3 out of 14 declarers make four hearts and 2 out of 8 declarers make three notrump.

Score	**IAF**
+420	+8.57
+400	+8.33
-50	-1.62
-100	-3.48

Contract	Lead	Tricks	Count	Overall	Bid	Play
4♡, E	♢	8	1	-3.48	+0.44	-3.92
		9	4	-1.62	+0.44	-2.06
		10	2	+8.57	+0.44	+8.13
	♠	9	2	-1.62	+0.44	-2.06
4♡, W	♠	9	3	-1.62	+0.05	-1.67
	♣	10	1	+8.57	+0.05	+8.52
		8	1	-3.48	+0.05	-3.52
3NT, E	♣	9	1	+8.33	-0.52	+8.86
		7	6	-3.48	-0.52	-2.95
	♠	9	1	+8.33	-0.52	+8.86

IAF scoring, declarer's view in Example 5

Contract	Lead	Tricks	Count	Overall	Bid	Lead	Other def.
4♡, E	♢	8	1	+3.48	-0.44	-0.59	+4.50
		9	4	+1.62	-0.44	-0.59	+2.65
		10	2	-8.57	-0.44	-0.59	-7.54
	♠	9	2	+1.62	-0.44	+2.06	0.00
4♡, W	♠	9	3	+1.62	-0.05	+1.67	0.00
	♣	10	1	-8.57	-0.05	-2.50	-6.02
		8	1	+3.48	-0.05	-2.50	+6.02
3NT, E	♣	9	1	-8.33	+0.52	+1.27	-10.12
		7	6	+3.48	+0.52	+1.27	+1.69
	♠	9	1	-8.33	+0.52	-8.86	0.00

IAF scoring, defenders' view in Example 5

We see that very little of the variation is assigned to the bidding. The table format makes clear that there is a single bidding score for the same contract played from the same side. The Valet score also likes the spade lead against four hearts from either side, which makes sense as no declarer had any chance after that.

Against four hearts by East, it was better to lead a spade (which always beat the contract) than to lead a diamond (which sometimes let declarer make). The lead scores correctly reflect this. When the defenders did subsequently beat the contract on a diamond lead, their defense score benefits.

Similarly, a spade lead against three notrump let the contract make, so that causes a terrible lead score.

So even this supposedly dull hand actually had some interesting aspects. To be sure, we could construct a hand that really did only depend on a single finesse. We could make it so that exactly one half of the declarers bid game and the other half didn't.

If the finesse is onside, we would then find that the game bidders would get a plus score for their bidding, and non-game bidders would get a negative score. There would be no variation in the play or defense. For a non-vulnerable three notrump, it would of course be +400 vs. +150. If there are a lot of pairs, then the game bidders would get approximately +3.00 IAF IMPs (as 250 corresponds to 6 IMPs) and the non-game bidders would get -3.00 IAF IMPs.

If we change the same hand by moving the finesseable card, then the plus-minus relations would change. The game bidders would get -50 and the non-game bidders would get +120. The IAF IMP scores would be -2.50 IMPs for the game bidders (as -170 is -5 IMPs) and 2.50 IMPs, respectively.

If all the pairs played a tournament consisting only of these two hands, and the same pairs bid game on both hands, then the game bidders would get a positive score of 0.25 IMPs per board (3.00 – 2.50 divided by 2), and the others would get -0.25 IMPs per board. This is as it should be. Luck evens out and skill remains in the long run. We might have wait a while, though.

North dealer
EW vul.

North:
♠ K 9 7
♡ 9 8 2
♢ A K 8 4 3
♣ 9 6

West:
♠ J T 3
♡ T
♢ Q T 6 5
♣ A K J 5 2

East:
♠ A Q 8 6 4 2
♡ Q 6 3
♢ 9
♣ Q T 3

South:
♠ 5
♡ A K J 7 5 4
♢ J 7 2
♣ 8 7 4

Our final example is from the 2015 Danish First Division. This one is interesting because declarers bid to different levels in the same denomination.

Contract	Declarer	Count	Tricks
5♢×	N	1	7
5♡×	S	1	8
4♡	S	1	8
5♠	E	1	11
4♠×	E	1	11
4♠	E	4 1	10 12
3♠	E	2	12

The overall IAF IMPs for North-South are as follows.

Score	IAF
-100	9.82
-230	7.45
-500	1.36
-620	-1.36
-650	-2.18
-680	-3.18
-800	-5.73
-990	-9.55

Contract	Tricks	Count	Overall	Bid	Play
5♢×, N	7	1	-5.73	-5.73	0.00
5♡×, N	8	1	+1.36	+1.36	0.00
4♡, S	8	1	+9.82	+9.82	0.00
5♠, E	11	1	+2.18	-3.45	+5.64
4♠×, E	11	1	+9.55	+8.18	+1.37
4♠, E	10 12	4 1	+1.36 +3.18	+2.15 +2.15	-0.79 +1.03
3♠, E	12	2	-7.45	-7.65	+0.20

IAF scoring, declarer's view in Example 6

Contract	Tricks	Count	Overall	Bid	Def.
5♢×, N	7	1	+5.73	+5.73	0.00
5♡×, N	8	1	-1.36	-1.36	0.00
4♡, S	8	1	-9.82	-9.82	0.00
5♠, E	11	1	-2.18	+3.45	-5.64
4♠×, E	11	1	-9.55	-8.18	-1.37
4♠, E	10 12	4 1	-1.36 -3.18	-2.15 -2.15	-+0.79 -1.03
3♠, E	12	2	+7.45	+7.65	-0.20

IAF scoring, defenders' view in Example 6

The solitary North declarer in diamonds gets his penalty in the bidding column. Similarly for the two declarers in hearts, as they played from different sides, so there is no comparison available.

The real interest is in the spade declarers. Thorvald was in 4♠ making 12 tricks. Why did he get +1.03 for his declarer play, while the declarer in 5♠ making 11 tricks got +5.64?

The reason is that the Valet score measures declarer play *relative to the actual contract*. Thorvald was only in 4, and his good play (or his benefit from the soft defense) didn't benefit his team a lot, as it was only the second overtrick. When we replace Thorvald by each of the spade declarers, we get somewhat lower IAF scores, but not a lot.

In contrast, the declarer in 5 really did have to take at least 11 tricks. In fact he did. If we replace him with each of the spade declarers, his team would have gone down five times out of nine, and each time it would cost a lot of IAF IMPs. Relative to this, his declarer play was very valuable. So he gets a much better play score for taking 11 tricks than Thorvald gets for taking 12. (He does get a poor bidding score though, and that is as it should be.)

2.5 ADJUSTMENT

Some people will never be content… Not satisfied with the false accuracy of all these Valet calculations, they also want to compensate for the strengths of the opponents faced by different pairs.

Since I was at it, I implemented such an optional adjustment in the Valet program. The adjustment calculates a weighted average of the overall Valet score of all the opponents faced. Furthermore, following an idea by Paul Gipson, the opponents' Valet scores are recalculated by excluding the boards played against the specific pair in question, on the theory that these boards would otherwise count "double". Of course it's not really double, as the boards generally have a small effect on the opponents' scores, at least in a big tournament.

In any case, the feature is there. I don't personally recommend using it except in rare cases where there are not so many different opponents, but then there will also not be so many comparisons, and the Valet score (and the Butler score) is of questionable utility.

2.6 INDIVIDUAL RESULTS

Some scoring programs output the Butler score for individuals, so if a player has played with multiple partners, a composite score arises. I don't think this is very appropriate except perhaps for declarer play and leads, so the Valet program does not offer this feature. It wouldn't be hard to implement, though.

3 Results

3.1 STRONG TOURNAMENTS

I've applied the Valet score to some strong tournaments whose results I could screen-scrape from Internet sites. If there were qualifying and final rounds, I only took the qualifying round-robin rounds.

- 2004 Olympiad (this was was done more or less manually from pdf files, though).
- The 2008 and 2012 World Mind Games, both for men and women. These tournaments replaced the Olympiad.
- The Bermuda Bowl and Venice Cup for 2005, 2007, 2009, 2011, 2013 and 2015.
- The European Championship for 2006, 2008, 2010, 2012 and 2014, both for men and women.
- The Polish first league for 2012, 2013, 2014 and 2015.

It was a fair amount of work, and in some cases there were small issues with data quality, but overall I think the data is quite good. From my own experience I would say that the pairs may well not have sat the way they were entered on the Bridgemates, so it's not a good idea to rely too much on individual as opposed to pair scores.

Due to the variations in Butler implementations, don't expect my results to be identical to the "official" ones.

The results are distributed together with the Valet program, including in pdf format. I have not attempted any statistical analysis on the number of hands needed for some kind of relevance, but I have often drawn the line at either 100 or 200 hands for a given pair.

I mentioned earlier that datum IMPs give rise to larger numbers than do IAF IMPs. I've quantified this, and the effect is 10%. So 10 datum IMPs are statistically similar to 9 IAF IMPs.

I've also quantified the relative effects of bidding, play and defense on results. This is a quick-and-dirty approach: I took the standard deviation of each of these three components across all pairs with enough hands played in a given tournament (regardless of how many hands each pair played). If there is more variation in the bidding results than in the defense results, that means that the bidding varies more and so plays a larger role.

Across all the tournaments, it turns out that the bidding counts for about as much as the play and defense together. This is calculated per hand, so if you play 100 hands, you might be expected to bid 100, declare 50 (25 for each of you and partner), and defend 50 hands. The stated results apply when dividing the IMP swings by 100 for each component (and not by 100 for bidding, 50 for play and 50 for defense). In other words, it is indeed a bidders' game.

3.2 CHEATING

I thought it might be possible to get indications of cheating by looking at some statistics across tournaments. In the Valet distribution I also include spreadsheets that cover all of the tournaments (except the Polish league games).

Some of the pairs that are reported to have been cheating do indeed score well on the Valet score, including on bidding, opening leads and defense (where cheating might be expected to help). But I think the signal-to-noise ratio in the data is not good enough to be helpful in this regard. Perhaps someone has a better idea for using this data.