A Response to:
Ian McKinnon's Article
On Victory Point Scales in Duplicate Bridge

As Published in January 2014 Issue of IBPA Bulletin

The January issue of the IBPA Bulletin provides a link to an article written by Ian McKinnon called *On Victory Point Scales in Duplicate Bridge*. This article, which is critical of the new WBF scale, is described by the IBPA editor as "a learned treatise." The article's author is widely recognized as both a highly respected writer and an innovator in all bridge scoring matters. However, as this lengthy riposte will demonstrate, while much of the article supports the approach adopted by the WBF Scoring Panel, it also contains several errors, generally lacks conviction where critical, and worse, offers some unacceptable alternatives.

Detailed responses to McKinnon's main points of criticism are given below and where possible technical issues are avoided.

1 Ties are not broken in a fair way

McKinnon contemplates the unlikely scenario where team–A is 0.01 VPs behind team–B at the end of the round robin. He shows that if team–A had won two previous matches by 1 Imp less in each, while team–B had also won by 1 Imp less in (perhaps) two other rounds, then the finishing order would be reversed. That is, team–A would have finished ahead of team–B by 0.01 VPs.

This is a specious argument at best. Consider a parallel situation in a soccer tournament where team–B wins the round robin by one point ahead of team–A. In soccer league tables, the winning team scores three points, the losing team gets zero points; if the teams draw they each get one point. Points on the league table are equivalent to VP's in bridge; goals are equivalent to our Imps. Now suppose in one of the round robin matches, team–A had won by two goals, while team–B had won by one goal. If, in that round, each team had scored one goal less, then team–A would have finished one point ahead of team-B on the league table. Who would argue, as McKinnon did in the bridge case, that team-A should feel aggrieved? In both cases, soccer and bridge, the conversion to VPs is a non-linear process, so that scenarios like those above can always occur. Team–A might feel unlucky, but has no basis to feel aggrieved due to unfairness.

In soccer tournaments, as in many other sports, the league (read VP) table is based solely on a win-draw-loss system. This is a particularly brutal scale which gives no credit to the loser for the size of its loss. Bridge scoring offers a

more sympathetic scale giving the losing team a proportionate number of VPs depending on how badly it loses. One might contend that goals in soccer are more important than Imps in bridge. That is possibly true, but the same situation occurs in cricket (and other) round robin matches, where it is arguable that a single run is less important than a single Imp in bridge.

What people have to get their heads around is that under the new scale not all Imps are equal, just as not all soccer goals or cricket runs are equal. Indeed, all extra goals/runs beyond the first winning goal/run are worth precisely nothing in soccer/cricket. At least in bridge, we give some credit for Imp margins beyond the first.

Because of the possibility of such scenarios, McKinnon would have total scores of less than one VP broken by some tie-breaking method. That is not a view that would be shared by most players or even most administrators. The chance of an exact tie on the new WBF scale is quite small so no tie-breaking methods are likely to be required. Indeed most observers have applauded this aspect of the new scale, rather than criticize it.

2 The blitz point should be reduced

To avoid scoring blow-outs, which can have serious repercussions for round robin results, the VP scale is truncated at a maximum Imp score. This maximum Imp score is called the *blitz point* and the value chosen is an important parameter of any VP scale. The new WBF scale uses a blitz point that is statistically determined from a large database of recorded results. These data indicate a standard deviation of about 7.5 Imps/board, a result which is consistent with other studies of Imp distributions. The blitz point in a match is defined by the WBF Panel as two standard deviations for the given number of boards in play. For the case of 12, 16 and 20 boards, the respective blitz points are: 52, 60 and 68 Imps respectively.

McKinnon suggests that the blitz point should be one standard deviation rather than two standard deviations; that is half the Imps listed above. No cogent reasons are offered for this radical suggestion. It is well-known that roughly 68% of all results will fall within one standard deviation, while 95% of all results will fall within two standard deviations. McKinnon's suggestion of a one standard deviation blitz would mean that nearly one-third of all matches would result in a maximum win of 20 VPs. Clearly that is an unacceptable outcome on any rational reckoning. Using a 20 point VP scale with twenty one-VP bins: 0–1 VP, 1–2 VPs, 2–3 VPs, $\cdots$, 19–20 VPs, roughly 5% of all results might be expected to fall within each bin. Indeed, this desirable feature of any VP scale, has been well verified for the major tournaments to which the new scale has been applied.

3 The new WBF scale is not strictly continuous

The new WBF scale is referred to as a "continuous" VP scale. In fact, it can only be a pseudo-continuous scale at best because the underlying Imps are always discrete integers. The WBF Scoring Panel used the term "continuous" mainly to distinguish it from the old "discrete" scales, as well as the fact that the new scale is based on a truly continuous mapping of Imps to VPs.

McKinnon complains that the WBF have not achieved the goal of continuous scales. What he means by this statements is somewhat technical and is described next, but in any case while this may be his own personal goal, it was not one of the Panel's.

The new WBF scale is controlled by two important parameters called the *sensitivity* and the *concavity* of the scale. The sensitivity is the value of each VP per marginal Imp[1]. The concavity measures the degree of non-linearity of the scale. Two overarching rules adopted by the WBF are as follows: for each marginal Imp up to the blitz point,

1. the sensitivity should be strictly positive, while
2. the concavity should be strictly negative.

Together, these rules imply that each marginal Imp is worth a positive VP increment, and this increment decreases as the number of marginal Imps go up. The concept is exactly analogous to the *Law of Diminishing Marginal Utility* in economic theory.

If adherence to these rules were rigorously applied, it would be necessary to present the VP scales to three decimal places. However, if the rules were relaxed just a little so that the concavity condition permitted zero as well as negative values, then the VP scale could be presented to two decimal places. Indeed, it would even be possible to present the scale to one decimal point, but in that case the scale would have zero concavity for almost every Imp margin. The choice of two decimal points is a compromise between the virtually linear one decimal scale and the "fully continuous" three decimal scale. The two decimal scale is found to have a relatively smooth degree of non-linearity, while the one decimal scale has significant jumps in sensitivity. Further, since the difference between the one and two decimal scales is very much greater than between the two and three decimal scales, the two decimal scale is a much better and closer fit to the ideal.

McKinnon, who agrees with the two stated design rules, argues that this compromise to two decimals has failed to achieve the WBF's goals of "perfect continuity" and asks why then the WBF even bothered to present its scale. This

[1]The marginal Imp score in a match is the difference between the winning teams Imps and the losing teams Imps.

contention is pedantic in the extreme – especially considering that practicality rather than perfection *per se* was also an important consideration of the Panel's design philosophy. McKinnon goes on to present a 16-board, 30 VP scale with one standard deviation blitz point. He claims this scale strictly satisfies both design rules above, but a close inspection shows that the concavity rule is violated in several places. Of course, if it were possible to obtain a one decimal scale satisfying both rules rigorously in a 20 VP scale, then that would obviously be preferred. However, such a scale is not possible, even when the blitz point is halved and the VP maximum increased as attempted by McKinnon.

4 No good reason for using a 20 VP scale

McKinnon further contends that there is no good reason to adhere to a 20 VP scale and that increasing the VP maximum would solve a lot of problems. Other VP scales eg. 0 to 30; 0 to 50 or 0 to 100 could have been used rather than the 0 to 20 VP scale offered by the WBF. Indeed, some of these scales might have slightly better sensitivity/concavity discrimination properties, but there is a good reason for sticking with the 0 to 20 VP scale.

First, the bridge public are used to VP scales of this magnitude. Early VP scales of the 1960's were 0 to 20 scales and the discrete WBF 0 to 25 scales were essentially minus 5 to 20 scales with 5 VPs added to remove the negative VPs. But more importantly, as described earlier, the choice of a 0 to 20 VP scale dovetails nicely with the two standard deviation blitz point to produce a scale that generates in practice, roughly equal numbers of outcomes in each one-VP bin. Finally, the WBF Scoring Panel have tested the new 0 to 20 VP scale on all board numbers from four up to forty without observing any adverse effects. Having a single maximum VP score for all practical board lengths is obviously more desirable than a system with variable VP maxima. Unless precise benefits for a higher VP maximum can be demonstrated there is no reason to consider it further.

5 The VP granularity is too fine

It is easy to make the mistake that a two decimal scale is not warranted by the nature of the game of bridge – in particular, that the second decimal could not possibly be significant given the high degree of random variation in scores. This argument might carry some weight if every possible score using two decimals between 0.00 and 20.00 VPs were permitted. But the fact is that only a small percentage of such scores are actually available. For example, in a 16-board match there are only 121 different VP scores allowed out of a potentially 2001 scores. This density of scores, or granularity of the scale, amounts to only six percent.

Thus while it is necessary to keep two decimals to generate a smooth non-linear scale (an argument, incidentally, that also applies to linear scales), this does not imply we are actually resolving match outcomes to that level of precision. In fact, the granularity of the scale is quite coarse rather than the opposite as claimed by McKinnon. It is feasible to multiply the two decimal scale across the board by a factor of one hundred so that a maximum win would be 2000 VPs. While this removes all decimal points, it does not change the granularity of the scale nor any of the essential features of the scale. In other words it amounts to little more than window dressing to placate the decimal point detractors.

6 The new scale changes bidding tactics

Any non-linearity in a VP scale will change the break-even probabilities of certain bidding strategies such as bidding a part-score, game or slam. The WBF Panel have looked at several different bidding scenarios to see the effect of this non-linearity. It transpires that the biggest effect is due to the blitz cut-off rather than the non-linearity in the scale. This large effect was already present in the old discrete scales because they also had a blitz cut-off. In theory, if a team discerned correctly that it was already down by a large number of Imps, then the mathematics dictates they bid their game/slam very aggressively; their opponents should do the opposite and bid very conservatively. There is nothing new in this tactic: when down a lot, the losing team might as well throw caution to the wind. There is no evidence to suggest that the sliding VPs (from 0 to 5) in the old discrete WBF scale did much to either deter or otherwise this normal behavior.

The concavity in the scale introduces a small change in percentage actions (relative to straight Imp scoring) amounting to some four percent for a vulnerable game and about seven percent for a non-vulnerable slam. However, as one can never be sure of the current state of the match, regardless how bad it may seem to be going, nor what one's opponents will bid on the hand in question, there will always be uncertainty risk to contend with. As a result, players were perhaps reluctant to apply these strategies in practice. It remains to be seen if they will do so under the new scale. One advantage of the new scales is that these sorts of calculations are much easier to make than with the old discrete scales.

So when McKinnon states: " the new scales should be looked at more carefully for both game and slam bidding", it might be noted that the Panel have indeed done so but it is still an open question how the players will respond to it. Most likely it will have little effect.

7 Conclusions

McKinnon concludes:

1. Even making those changes[2] does not address the ease of use and simplicity concerns expressed by the bridge playing public.

2. Unless changes are made to produce a truly continuous scale there does not appear to be any advantages in (the WBF) scales, and in fact there are many disadvantages.

3. I would like to suggest that other choices in the sensitivity model, a (blitz) cut-off point between 1 and 2 standard deviations and a different choice of maximum VP scales may solve some or all of the problems in achieving the current WBF objectives.

4. If the bias in the current (old) scales are meeting their needs then the complex, partially continuous scales are not needed and are unfair. A simpler approach such as suggested by Ron Klinger in the July 2013 ABF Newsletter (and IBPA Bulletin) are clearly sufficient.

The IBPA editor adds his own concerns:

5. The debate over the new WBF VP scales rages on. That the scale produced such debate in the first place is an indication that all is not quite right with it.

6. The WBF has done as it promised ... and conducted a survey among interested parties. We eagerly await the results of the survey questionnaire.

7. Our chief objections to the new scale are that (i) it is too complex for easy reference, and (ii) it is completely unnecessary. Does it not seem awkward (at best) when a team scores up a match and does not know the VP score without reference to what appears to be a logarithm table!

8. Every other sports league that we know of, including chess, uses win/loss/draw as a method of determining its winners and/or qualifiers.

While the Panel welcomes debate on the new scale, much of the criticism above is uninformed and in some instances plain wrong. It gives little credit to the expertise and deliberations of the Panel and fails to acknowledge that similar versions of the scheme have been successfully used for some ten years in the USBF teams selection trials. Presented below are some measured responses to the eight concerns listed above.

1. Generally the bridge playing public have not expressed dissatisfaction with the new scale. Much of the criticism has come from viewgraph commentators and bridge journalists. What exactly is so difficult about looking up a

[2] McKinnon's proposed changes described in section 2 of this response

VP score for a given Imp margin in a table published inside a score-book? This was exactly the process used in previous incarnations of the WBF scales. The critics are perhaps confusing "ease of use" with "understanding the technical details that produce the scale." In practice, players need only worry about the former and not be concerned with the latter at all. Everyone is used to two-decimal currency; it's really no big deal to go to two-decimal VP scales. Indeed, many sports are now doing just that – changing to decimal scoring.

2. This statement is both unfounded and unfair. There is no practical need to pursue the goal of perfect continuity as espoused by McKinnon. The WBF has achieved the main targets it set out to do. Here are some of the advantages of the new scale: (i) it removes the cusp-effect by making every Imp count; (ii) it ensures that later marginal Imps earned in a match are never worth more than earlier marginal Imps; (iii) marginal Imps in a short match are always worth more than marginal Imps in a longer match; (iv) the chances of a tie in a long tournament is virtually zero; (v) it admits a VP distribution which is roughly uniform in each VP band; (vi) the same formula is used to generate the VP scale for every practical board number. Most of these features were absent in the old discrete scales.

3. McKinnon's suggestions and calculations for improving the scale are incorrect. The scale he proposes does not actually work and the idea of a one standard deviation blitz point would lead to an unacceptable number of maximum VP wins per round. McKinnon overstates the objectives that the new scale seeks to address. As affirmed above, the WBF targets for the new scale have been achieved.

4. The notion of unfairness of the scales has not been demonstrated by McKinnon. The one example he gives is easily refuted and is just a direct consequence of the non-linearity of the scale. The virtually zero chance of getting a draw over all matches in a round robin is a positive, not a negative feature of the new scale. If it is good enough to win a cricket match by one run, why is it not good enough to win a bridge match by one Imp? One run in cricket is just as random, if not more so than one Imp in bridge.

   The reference to Klinger's " simpler" linear or quasi-linear scale, while feasible at first blush, has a number of negative features. The concavity of these scales is zero (rather than negative) almost everywhere and they have significant jumps in sensitivity. While the quasi-linear scales can be presented to one decimal place, the cost of an extra decimal is compensated by a smoother sensitivity profile and better concavity discrimination. In technical terms, these one-decimal quasi-linear scales are really no simpler than the two decimal scales.

5. The conclusion that there must be something wrong with the scale because there has been largely negative debate, does not hold up if precise failings

of the system are not presented. The "what-if" scenarios of Klinger's article are all consequences of the non-linearity of the scale. If a critic can only think linearly, they will always find something wrong with the scale, but others will accept the consequences that non-linearity implies. For example consider as Klinger did, that a team wins 50 Imps over 10 matches. On a linear scale, the team's VP score will be the same regardless how the 50 Imps are distributed. On the new WBF scale it matters how these 50 Imps were accumulated. The highest VP score will occur if the team wins all its matches by 5 Imps and lowest (barring losses) if it wins one match by 50 Imps and draws the other nine. This difference is a direct result of the non-linearity of the scale and makes perfect sense if emphasis is placed on winning matches.

In any case, since the non-linearity is quite small, these consequences are by no means extreme. The failings expounded by McKinnon are all related to his view, not shared by the WBF Panel, that the scale should be perfectly continuous. The Panel have consciously and deliberately chosen a more pragmatic approach that such perfection is impractical, nor is it necessary.

6. The results of the survey on the new WBF Scale have now come to hand and comprehensively show an overall approval of the new scale and a strong preference compared to the old discrete scales. It would appear that contrary to the negative protestations heard in some quarters, the majority of players have given the scale the thumbs-up. While the mechanics of the scale are not yet generally understood (perhaps most respondents have not read the WBF's report on its website), this will change with time and usage.

7. The "too complex for easy reference" comment seems to be a popular journalist's catch-phrase. As mentioned in 1. above: players, directors, administrators and journalists need only look up a table to convert an Imp score to a VP score – hardly too complex or onerous a task. Most people will be uninterested in the mathematics that generate the scales. In truth, for the technically adroit bridge public, the mathematics of the scheme is not that complex anyway. The scales for any board number are generated by a single formula – which is a walk in the park for any computer and easy to understand by anyone competent in high school algebra.

   The editor's statement that the scale is completely unnecessary is not tenable. There are very good reasons for converting Imps to VPs in round robin teams matches. It is basically invalid to add Imps from one round to another without first standardizing the results. The reasons for this are well-known: some board sets may be quite flat, while others wild and variable. Different opponents from one round to the next and the randomness in decision making introduce further variability. Ideally we

should standardize the mean and variance of the results of all matches within a round before adding the scores. This would entail submitting the results and waiting for the computer to do the analysis before getting each team's score. No-one would really want to follow that path, so the use of a VP scale at least approximates the standardization process, and permits an instant result for the match. The editor's argument that the new scale is unnecessary because it produced *almost* the same outcome as other simpler scales on the Bali results, should ring alarm bells. Quite apart from the fallacy of a one tournament comparison, some of these scales produced the same qualifiers but in a different order; others different qualifiers. These differences are significant and cannot be glossed over by a dismissive irrelevance of the new scale.

The editor would have us believe that some people really do consider that two decimal places is beyond their comprehension if his reference to logarithm tables is to be taken seriously. The consensus however is that bridge players have a lot more intelligence than he gives them credit for. They will with time and familiarity be accepting of the new scales.

8. The idea of a win/loss/draw scale for teams bridge would be a great tragedy for the game. Bridge is a unique sport in several significant ways. One is that bridge allows players with average skill levels to play against national champions. If the VP scale were a simple 2–0–1 (win/loss/draw) what chance would the average player have of winning? Virtually none. Yet, with a 20 VP scale they may well earn some reward for the occasional deals they outplay their more illustrious opponents. A win/loss/draw scale, while meritoriously comprehensible to all, might not only drive away new adherents to the game, a charge leveled against the new scale, but established players as well.

Much of the published criticisms and concerns of the new scale are largely uninformed. Some of the reactions are more of the knee-jerk variety rather than of rational debate. A personal anathema to decimal points is not a rational basis for criticism, particularly when their existence is confused with complexity. As has been discussed in these pages, two decimals are the optimal needed under the carefully reasoned design constraints imposed.

While McKinnon's contributions to the bridge world[3] have been rightly admired and rewarded, his article, which has prompted this response, cannot be construed as an enlightened piece let alone a "learned treatise". In particular, his wish for perfect continuity can only be obtained by going to three decimals, even with some relaxation of the design parameters. His goal of achieving it with just one decimal is basically unattainable.

[3]In terms of movements and scoring.

One situation where it is possible to appease the two decimal haters, is to display the current VPs in a viewgraph presentation or bulletin article to only one decimal, or to the nearest 0.5 VP or even no decimals at all. The first of these (one decimal) was trialed in the 2013 Bali championships for all the teams matches[4] being displayed in the viewgraph room. This seemed to meet the approval of most participants and was certainly easier on the viewer's eyes. Naturally, the missing second decimal was kept in computer memory to ensure integrity of the scoring system.

It is the view of the WBF Scoring Panel that the new two decimal scale is based on sound principles and that the underlying mathematics and statistics have been correctly applied. The panel accepts that there will always be criticism of the scale because of its perceived complexity in some quarters. But this complexity is really an illusion – there is inherently no difficulty from the players' perspective in obtaining a match result from an Imp to VP conversion table. There are good reasons for computing the VPs to two decimals as described above. The players, administrators, commentators, journalists and any other stake-holders should feel confident that the new scale has been carefully and professionally worked out. Its implementation has many advantages over the old discrete scale and the WBF Panel are confident that the scale is not unfair as asserted in the recent McKinnon and earlier Klinger articles.

[4]Bermuda Bowl, Venice Cup and D'Orsi Trophy