

To Whom Tribute is Due:

The Next Step in Scoring Systems

Brandon Fogel, March 2020

All Diplomacy players agree that the primary objective of the game is to solo—gain control of at least 18 supply centers. There is not widespread agreement on secondary objectives, however, i.e., what one should aim for if a solo is not possible. Because most games do not end in a solo, this is a serious problem. If the secondary objectives are not well-defined and they are needed regularly, then the game is not well-defined enough for typical play.

Scoring systems were invented to solve this problem. By placing point values on various outcomes, scoring systems define which ones are worth pursuing and in what proportion, i.e., what the secondary objectives are. It is commonly thought that scoring systems are needed only for tournament or league play, but they are really an integral part of the game's definition. Different scoring systems define different versions of Diplomacy, with significantly different incentive structures that can result in dramatically different styles of play.

Consider how people play the two most common scoring systems, draw-size and sum-of-squares. In draw-size, there is strong incentive to find a good ally quickly and stick with the alliance the entire game. In sum-of-squares, there is significant incentive to break an alliance in the mid-game and go for a large center count. The most consequential decision in Diplomacy, whether to cooperate or try to dominate, depends on the scoring system.

Unfortunately, there is no consensus on a fundamental or primary form of Diplomacy. Some in the hobby think that a solo should be considered the only result of value. Others think that ending the game with the most centers is a good result, and the more the better. This leads to disagreement over what it means to “win” a game of Diplomacy. As I discuss in Part 1, even the published rulebooks (all 7 of them) do not speak univocally on the matter.

Debates over scoring systems can be highly subjective, largely because we lack objective means of comparison. Yet the hobby has shown clear preference for certain systems over others; sum-of-squares has replaced draw-size as the most popular system in the North American hobby because many players think it promotes a more exciting and rewarding style of play. Thus the various opinions about scoring are not merely arbitrary or whimsical, despite the vagueness that has plagued the debate about their merits.

In Part 2, I offer some objective analytical tools for comparing scoring systems. I do this by identifying certain general incentives that are desirable or that are widely valued throughout the hobby (or both) and then providing quantitative, combinatoric interpretations of them. There is still subjectivity in the choice of incentives, but I think there may be reasonable consensus on those. Nearly everyone values staying alive and acquiring supply centers, for example. In any case, debating the value of general incentives allows for a deeper, more sophisticated discussion than debating the scoring systems directly.

I conclude in Part 3 by introducing Tribute, a new scoring system that follows somewhat naturally from a straightforward implementation of the chosen incentive measures. There may be no perfect scoring system, but I believe Tribute is a step forward, offering a unique balance of incentives with an emphasis on dynamic gameplay.

1. To Win or Not to Win

1.1.A Little History

Starting an argument among Diplomacy players is not a difficult thing to do. One surefire way is to ask them to define “win”. For many, a win is a solo and nothing else. For others, a board-top is enough. For the solo purists, a board-top is merely a draw in which everyone either wins or loses equally (the choice depending on the level of misanthropy in the room). The debate can quickly take on a religious tone because there is no definitive way to resolve the question.

Those hoping for a textual resolution to the “win” question will be disappointed. The various published rulebooks are notoriously and even hilariously opaque on the matter. Calhamer’s original self-published text (1959) says that whoever gets a majority of pieces (not supply centers) is the winner. Absent anyone achieving a majority, the game is a draw, he writes, without indicating whether that should constitute a shared win. However, he is perfectly clear on who deserves to be shamed: “Those losing all their pieces lose in any case.”

The first Games Research Inc. rulebook (1961) removes mention of a draw, advises players to set a time limit for a “short game”, and then stipulates, “the player with the most pieces on the board at that time is the winner.” The second Games Research rulebook (1971) switches the criterion to supply centers, not pieces, and says that players may agree to end a game before anyone controls 18 of them, in which case all surviving players “share equally in a draw.” Separately, if a previously agreed-upon time limit is reached, the players “may agree to regard the player who has the most pieces on the board at that time as the winner.”

The text remains unchanged through all Avalon Hill editions until the most recent (2000), where the language about time-limited games is simply omitted. The wonderfully obtuse “share equally in a draw” survives.

Of course, the hobby has long ago left the cradle offered by the board game publishers, so even if the rulebooks were clear and univocal, it wouldn't matter much. For years, the various Diplomacy communities have experimented with different variations of the rules and found reasons to prefer certain elements over others. The differences are most prominent with win conditions, but there are others (e.g., whether draw votes should be anonymous, which no official rulebook addresses).

1.2. Value and Scoring

The specification of win conditions has dramatic consequences on gameplay. Players pursue very different strategies to top the board versus merely surviving to a draw, and the entertainment value of the game varies considerably as a result.

The typical way to specify win conditions is to assign point values to each possible result. Such a set of rules is known as a scoring system, although it might be better referred to as a system of values, since the system defines what results are valuable and in what proportion. Value provides incentive to select certain strategies over others. Without a complete system of values, the game of Diplomacy is not well-defined, since the system of values defines what the players should be aiming for.

A common misconception is that a scoring system is only needed in tournament settings, when player performance is being compared across multiple games. Scoring systems do facilitate this, but the need for a system of values is fundamental to the game, even so-called “house games” (one-off games not part of a tournament or league). People approaching a house game with the idea that a solo is the only worthwhile result are employing a specific system of values, which we can refer to as “solo-or-bust”. Specifying that the house game is to be played under draw-size scoring or sum-of-squares is perfectly reasonable and will result in different styles of play. For this reason, the scoring system should be made explicit, even if it is solo-or-bust. If different players were to play under different systems of value, they would not actually be competing against one another, at least not in a meaningful sense.

Imagine a game of Scrabble in which one player thinks that in order to win one must have the most points *and* score above 400. Another thinks only that one must play the Q, X, and Z in order to win. The mechanics of their two versions are the same, drawing tiles and making words, so they can go through the motions of playing together. They may even think they are playing the same game. But they aren't, and at some point the difference will manifest in some unpleasant way. Perhaps one player thinks the game is over once the Q and X have been played by different players, or the other claims the game is a draw even after being outscored 350 to 200. The result is a failure to engage in

meaningful strategic competition, because there was not agreement on how possible game results should be valued.

The analogy with Diplomacy under different systems of value is precise. In order to play a meaningful strategic competition, there must be an agreed-upon system of values. If one person thinks a board-top is important and another doesn't, they aren't playing the same game, even if they are both negotiating and writing valid orders.

Why isn't this a pressing problem in Scrabble? The official win condition—have the most points once all the tiles are in play—is straightforward and nearly always results in a victory for a single player. Not so with Diplomacy as originally conceived. Calhamer's idea was that the natural win condition is total domination, and that every game would achieve this result if played long enough. The majority-of-pieces condition was meant to reflect a tipping-point, after which total domination was inevitable. One can quibble about whether majority-of-pieces is a good proxy for total domination, but that is independent of whether total domination is what *should* be valued. What matters is that for Calhamer, and probably most early players, a game failing to end in total domination was only a matter of inconvenience, a consequence of players living lives outside of the game. Most of today's players don't think about Diplomacy this way, although we do accept that most games get cut short of a natural ending point (without the game being ruined for it). Scrabble would not be able to get away with this.

2. Incentives as Analytical Tools

2.1. The primary incentives

To answer the question of what *should* be valued in Diplomacy, I will offer a mix of subjective and objective considerations. My views about what makes a good strategy game are subjective, of course, although I believe most enthusiasts would approve. The selection of incentives to focus on is also subjective, although I have tried to select those that I believe are valued collectively by the hobby. The quantitative measures of incentive in the different systems are objective.

A good strategy game forces players to make difficult decisions between competing strategies. Diplomacy provides a wider set of viable strategies than most games, and this is one reason it is almost endlessly playable. Systems of values, or scoring systems, can promote certain strategies over others, sometimes strongly, with considerable effects on the difficulty of the game.

In order to evaluate the relative merits of different scoring systems, it is useful to look at the incentives they promote. To do this, we must first specify the incentives worth paying attention to. My goal is to identify incentives that are widely valued throughout the hobby as well as those that promote challenging and exciting gameplay, in accordance

with the views about strategy games articulated above. I have thus chosen the following incentives:

1. **Board-Top Incentive:** How valuable is having the most supply centers?
2. **Survival Incentive:** How valuable is avoiding elimination? How much do small powers have to play for?
3. **Growth Incentive:** How valuable is gaining more supply centers?
4. **Dominance Incentive:** How valuable is continued growth after taking the lead?
5. **Balance of Power Incentive:** How much more valuable is fighting the leader than fighting other powers?

The board-top, survival, and growth incentives are natural. Supply centers are the only elements of intrinsic value in Diplomacy, so acquiring and protecting them must be a core part of the goal of the game, however that is understood. And whether the scoring system in use values finishing with the most centers, most people appreciate doing so, even those who value the solo above all else. Likewise with survival; even with sum-of-squares scoring, where finishing with 1 or 2 centers has almost no value, players are still glad not to be eliminated.

The dominance incentive can be similarly justified. A football game won by a score of 49-14 is generally more impressive than 28-27. A Scrabble game won 300-100 is generally more impressive than one won 250-240. The interpretation of “dominance” in Diplomacy is less straightforward, since the competition is not binary and there is a zero-sum competition for the elements of value (supply centers). Generally players in the hobby are impressed by the overall size of a power rather than the margin of victory over the second-place player, although sum of squares scoring has encouraged attention to the “delta” between the top two powers.

The balance of power incentive is more esoteric but is perhaps the most interesting of all; I value it for its effect on gameplay, which is to promote second chances. If there is strong incentive for everyone to pull the leader back to the pack, then the game should be more dynamic, offering everyone a greater chance at succeeding even if they’ve fallen behind. In most scoring systems, smaller powers gain more by fighting each other than the larger powers, which actually helps the leader. This creates a snowball effect, leading to games that get quickly tracked into irreversible paths when one player gets an edge (e.g., Risk), which can reduce entertainment value. A strong balance of power incentive counters such snowballing and can promote dynamic games with dramatic changes of fortune.

The other incentives can also have desirable effects on gameplay. A strong board-top incentive makes unbreakable alliance play less appealing; alliances will buckle under the weight of their own success, as allies gain incentive to stab each other. Similarly, a strong dominance incentive should encourage more risk-taking and less playing-it-safe, which will lead to more spectacular (and entertaining) rises and falls. A good survival incentive gives smaller powers a continuing stake in the game, meaning less janissarying and metagame dot-throwing.

2.2. Quantifying the incentives

To facilitate the analysis of scoring systems in terms of incentives, I offer the following “next-dot” interpretations:

1. **Board-Top Incentive:** How much does taking the lead improve one’s score?
2. **Survival Incentive:** How much more does a 1-center power score than an elimination?
3. **Growth Incentive:** How much does taking a center increase one’s score?
4. **Dominance Incentive:** How much does taking another center improve the leader’s score?
5. **Balance of Power Incentive:** How much better is taking a center from the leader over taking one from the other powers?

These values can be calculated over the set of all possible changes on all possible relevant board configurations (i.e., the set of all supply center count distributions where the largest count is less than 18). Without good reason to do otherwise, I take each board configuration to be equally likely.

2.3. Existing scoring systems

These are the systems currently in widespread use:

- **Draw Size (DSS):** All surviving players split points equally.
- **Sum of Squares (SoS):** Players score in proportion to the square of their center count.
- **Carnage:** Players score in proportion to their rank plus a tiny bonus for center count.
- **C-Diplo:** Players score in proportion to their center count plus 1st, 2nd, and 3rd place players score a fixed bonus that is shared equally on ties.

In all of the analyses that follow, the total scores awarded in each scoring system are normalized to 100. The score of 7.1 (half of 1/7, the average pregame expectation for the average player) is marked as a measure of substantial score.

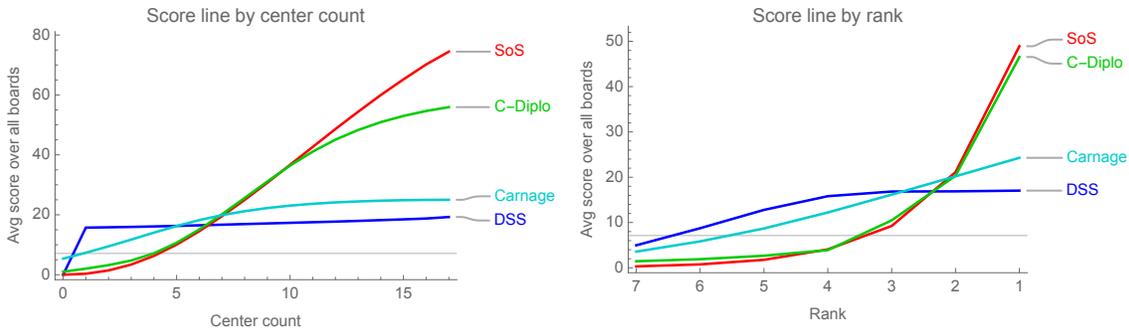


Figure 1: Average score over all possible boards as a function of center count and rank for the most common scoring systems. Survival and growth incentives can be read off the chart.

Figure 1 shows the average score over all possible non-solo boards, first as a function of center count and then of rank. A few things are evident from the score lines. Carnage and DSS are fairly flat, whereas SoS and C-Diplo regularly award big scores. Survival incentive can be read from the low end, especially the difference between center counts of 0 and 1. DSS gives the highest reward for survival, as expected, since survival is the only result of value in the system. Carnage awards points to small powers, but on average a 1-center power doesn't get much more than an eliminated power, so the survival incentive is fairly small (Carnage is alone among major systems in awarding a substantial amount of score to eliminated players). SoS and C-Diplo provide almost no survival incentive. This is a common criticism of SoS; once a power is pushed down to 3 centers or fewer in the midgame, there is little to play for, and players often turn quickly to janissarying or metagame considerations. The growth incentive can be read from the slope of the curve at any given point. (For charts of all possible scores for each system, not just the average score, see the Appendix.)

The score lines by rank show that SoS and C-Diplo usually award significantly more points to 2nd place than 3rd and below. This means that a “good 2nd place” is possible, which has the effect of encouraging alliance play.

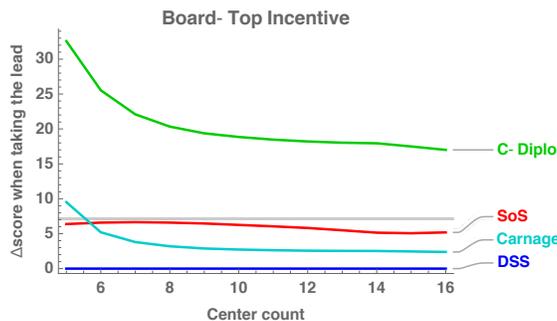


Figure 2: Average score change when taking the lead as a function of center count.

Board-top incentive is shown in Figure 2. C-Diplo provides strong incentive to take the lead, SoS provides some incentive, Carnage provides small incentive, and DSS provides none at all. This result may surprise proponents of SoS, since its principal feature is supposed to be that it encourages players to go for big center counts. But SoS also usually gives substantial reward to 2nd place, and this lowers the differential value of taking the lead. (One may wonder why the board-top incentive for C-Diplo is not 38; this is because jumping from 2nd to 1st is worth 24 points, while breaking a tie for 1st is only worth 12; the average for all boards is around 20.)

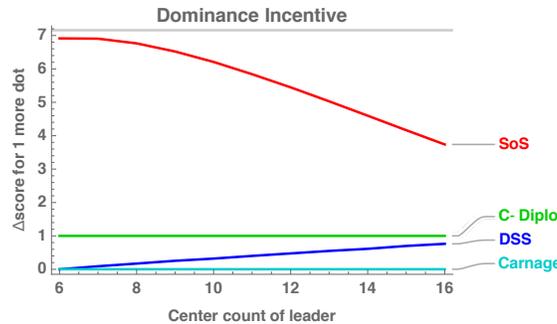


Figure 3: Average score change for the leader when taking a center from any other player.

SoS is the only system that provides any substantial dominance incentive (see Figure 3), although it decreases with overall center count, perhaps a surprising result. With C-Diplo, once a player has the lead, further centers are only worth 1 point. DSS provides a small incentive to continue growing, because doing so will sometimes mean eliminating another player. Carnage provides only miniscule scoring incentive for the leader to keep growing.

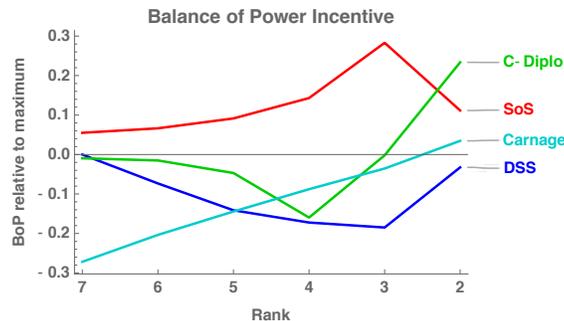


Figure 4: Average difference in score change when a player takes a center from the leader versus another player, as a fraction of the theoretical maximum, as a function of rank

The balance of power incentive (Figure 4) measures how much better it is for players not in the lead to take a center from the board leader rather than another player. The theoretical maximum for this differential is about 2.6 points (on average) for each rank except 2nd place, where it is about 15 points. To reduce skewing between these ranks, it is useful to look at these differentials as fractions of the theoretical maximum. What stands out on the chart is that most of the systems provide negative incentive for most players. In DSS, Carnage, and C-Diplo, it is almost always better for smaller powers to

fight each other rather than join against the leader. SoS fares a little better here but is still fairly weak, especially for the lower ranks.

While alliance play is too complex to admit a single incentive measure, the board-top, balance of power, and dominance incentives taken collectively may be a suitable proxy. SoS has a clear advantage over the other systems in these incentives, although C-Diplo gets notice for a large board-top incentive. Still, neither is strong on all three measures.

There are other considerations that are not easily quantified. A scoring system should be simple, easily understood and able to be calculated on the fly. DSS and Carnage do well here, C-Diplo does fairly well, and SoS does poorly.

Other features are often valued, although with less consensus. For example, DSS where surviving players can agree not to participate in the draw generally leads to shorter games than other systems, since alliance structures tend to remain static once established. Players typically agree by 1905-6 that the outcome is clear. Some players may consider such speed a virtue of DSS, others a drawback.

3. The Tribute Scoring System

Is it possible to construct a system that promotes all five of the incentives discussed in Part 2 while remaining fairly simple? The answer is yes, as will be shown next.

3.1. Implementing the incentives

To promote the board-top incentive, a system should award a bonus for topping the board, and the award should be substantially higher than any bonus awarded for a shared top. Awarding a bonus for 2nd or lower places will decrease the board-top incentive and should thus be avoided.

To promote survival incentive, a system should award a bonus for survival. Any points awarded to eliminated players will decrease survival incentive and should thus be avoided.

To promote growth incentive, a system should provide higher scores to players with more supply centers.

To promote dominance incentive, a system should provide a bonus to the board-topper that increases with the size of the power, the margin of victory, or both.

To promote balance of power incentive, a system should put the survival bonus in competition with the board-top bonus. The better the leader does, the worse the survivors do, and vice-versa.

3.2. The new system

To help locate a well-balanced implementation of these incentives with maximal simplicity, I enlisted the help of fellow Weasels, in particular Jake Trotta, Bryan Pravel, and Chris Kelly. Together we landed on the following system.

Games ending in a solo award 100 points to the soloist and 0 to the other players. For all other games:

- Each player gets 1 point per supply center (Growth)
- All survivors split 66 points equally (Survival)
- Every surviving player pays 1 point in tribute to the board-topper for every center he/she has over 6 (Board-top, dominance, balance of power)
 - A player cannot give more than his/her share of the survival pool
 - Shared toppers split the tribute equally

The name Tribute has been chosen for this scoring system to emphasize the payment survivors must make to the board-topper. This is the key dynamic aspect of the system; it forces all players to always have a stake in what the board leader is doing.

The choice to exempt the board-topper's first 6 centers from the tribute is due to the fact that 6 is the smallest possible center count that a board-topper can have. Thus the board-topper is only rewarded for performance over the minimum.

To generate sufficient survival incentive (on average at least half the pregame expectation value), the survival pool should be roughly twice the size of the center count pool. Since 100 is a nicer number than 102, we chose 66 rather than 68 for the survival pool. 60 would simplify the mathematics, but the convenience of having the total number of points add up to 100 is too great to pass up.

The net result of these choices is that most scores in Tribute can be easily calculated without a calculator. At the very least, it's easy to see how scores will change based on transfers of supply centers. If you take a center, you gain 1 point; if the center is from the board-topper, you gain 2 points; if you take the lead, you gain a lot more points (1 plus the number of players left in the game times your center count above 6, to be exact).

A sample calculation:

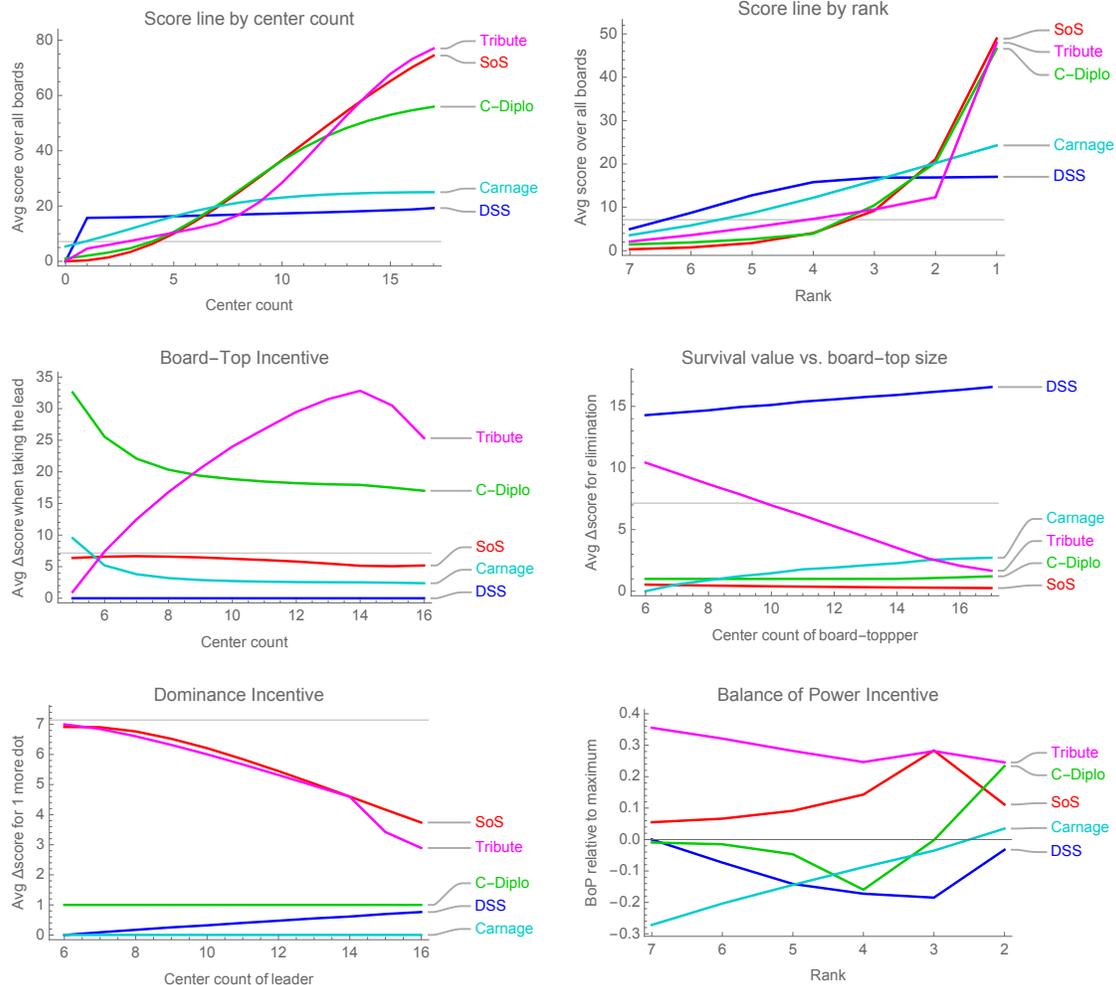
Power	Centers	Survival Bonus	Tribute	Total
Austria	9	11	-4	16
England	3	11	-4	10
France	0	0	0	0
Germany	10	11	+20	41
Italy	5	11	-4	12
Russia	4	11	-4	11
Turkey	3	11	-4	10

Tribute's incentive structure can be boiled down to these simple slogans: *Survive, grow as big as you can, top the board if possible, otherwise keep the board-topper as small as possible.*

For those who worry that adding a survival incentive will mean that players will focus on eliminating others, thereby leading to unpleasant gameplay (a common criticism of DSS), note that in Tribute there are two counterbalancing incentives: 1) the board-topper has incentive to keep smaller powers alive in order to collect more tribute, 2) non-toppers have to maintain focus on the board-topper in order to avoid paying less tribute.

3.3. Comparison to other systems

Tribute is compared to the major systems in the following charts. It does well in all 5 incentives.



With strong dominance and board-top incentives, and a good balance of power incentive, Tribute should discourage unbreakable alliance play. 2nd place generally scores significantly lower than in other systems, so the notion of a “good 2nd place” should be less enticing. With a decent survival incentive, Tribute should encourage smaller powers to stay engaged in the game. And with a good balance of power incentive, smaller powers have reason to focus on the bigger powers rather than each other, hopefully leading to more dynamic games with bigger reversals of fortune.

Tribute should provide all powers, small and large, with a bigger set of viable strategic options of varying risk and reward than other systems; I take this to be a hallmark of a good strategy game. It is our hope that, by emphasizing multiple competing incentives, especially board-top, survival, and balance of power, Tribute will help enhance Diplomacy both as vehicle for entertainment and as a measure of strategic and diplomatic skill.

Appendix – All Scores

