A Lens Model of Expert & Novice Spades Players

by

Bryan Campbell

Elaine Mahoney

Keenan May

for

PSYC 461/527

Mike Byrne

INTRODUCTION

In many domains, humans are asked to make judgments. In the business world, people are asked to judge how well a prospective employee will perform in a specific job. In the medical world, doctors are asked to judge the likelihood of a certain diagnosis based on symptoms. In the meteorological world, weather forecasters are asked to judge the likelihood of certain types of weather. The examples are vast and span many fields, and as such, understand and predicting how people make these judgments became of great interest to psychologists. The lens model was developed from the work of Brunswick (1952) and Hammond (1975) as a tool for investigating and evaluating people's judgments. This model assumes that people make judgments based on the values of certain cues that are available to them at the time they are asked to judge, and that a given judgment scenario can usually be represented as a linear summation of a relatively small number of cues. The cues themselves are pieces of information that take on or can be assigned a value on some numerical scale. The relationship between these cues and a person's judgments are probabilistically related; thus, by using multiple regression, one can create a statistical model of the judge.

Once built, a lens model presents a large amount of information. By examining the multiple regression equation for a judge, one can learn about his/her cue utilization, or which of the investigated cues the judge is actual using, as well as the weight placed on each of them. A lens model can also provide a measure of the judge's cognitive control, or how well a judge follows his own model. This value is the correlation between what the model of the judge predicts and the actual judgments made. Low cognitive control may indicate that a judge is inconsistent or that there are cues that the judge is using that have not been considered in the model.

When the true outcomes are known (such as actual performance in a job or weather results) another side of the lens model can be created. Like the model calculated for the judge(s), a model can be created that describes the environment in which the judgements are being made. In calculating this model, a measure analogous to cognitive control can be estimated for the environment. This measure, known as predictability, is the correlation between actual states and what the environmental model predict would predict. Low predictability, like cognitive control, can mean two things: a noisy, unpredictable environment or missing cues in the model.

Once models have been created for both a judge and the environment, the model can begin to show how well the judge does at predicting the environment. This evaluation is calculated with two different measures: achievement and knowledge. Achievement is the correlation between the judge's actual judgments and the actual outcomes, while knowledge is the correlation between the model of the judge's predicted judgments and the model of the environment's predicted outcomes. Both of these values are important for the evaluation of the judge's accuracy because it is possible for a judge to have high knowledge and low achievement. This pattern of results would indicate that the judge is aware of how the environment actually works, but because of issues of environmental noise or cognitive control, cannot make judgments that match the true outcomes.

In many cases, the environmental side of the model cannot be known due to a lack of criteria for what is "correct". That does not, however, mean that the model does not have value. As mentioned earlier, with only a single set of judgments, one can learn about a judge's cue utilization and cognitive control. Further, lens modeling can be used to compare the judgments of two different judges. In this type of model, multiple regression equations are calculated for each of the judges and cue utilization and cognitive control measures can be analyzed for both of them individually and compared across the judges. Two additional measures can also be calculated: policy agreement and agreement. Policy agreement is the correlation between each of the judge's model's predicted judgments and it is a measure of how similarly the judges use the cues that are being modeled. Agreement is the correlation between each of the judge's actual judgments. This measure tells us how much the judges agree when given the same values on all of the cues. Like the relationship between knowledge and achievement, it is possible for two judges to have low agreement but high policy agreement due to issues with cognitive control.

Lens modeling is more applicable to some domains more than others. It is especially good at revealing the cues used in scenarios where objective feedback is prompt and possible and decisions must be made frequently and under an amount of uncertainty. We were interested in applying the lens model to situations in which adversarial or cooperative game theory was used. The card game Spades provides the perfect environment for such a study. As such, a goal of the present study was to better understand how expert and novice players used literal cues, such as hand valuation, in the betting process in the game of Spades. This exact domain, and indeed this sort of execution-based decision-making in general, have not been studied at length.

The Game of Spades. According to Rayment (2011), Spades was invented in the 1930s, and became popular among the US military during World War II due to its relative simplicity, modularity, and portability (it doesn't require chips, and can be played in short bursts or at length). The game has an understandable property to it similar to Blackjack in the sense that players can become skilled at predicting their own performance, and that of the other players. It is played with two teams of two players. Each player is dealt 13 cards, and then asked to bet how many hands, or 'tricks' they will win using those cards. Bets proceed sequentially starting with the player to the left of the dealer (so players have progressively more information as the betting goes around the table).

Hands proceed by each player laying down a single card. The winner of the hand is the player that plays the highest card in the suite that was initially played, or who plays the highest spade, which is the trump suite. It is important to note that spades must first be 'broken' by a player who doesn't have any cards in the desired suite laying one down. After this, they may be played at any time. A winning strategy in the game of Spades is not necessarily about 'winning' the most hands—it is more about being able to predict what you will do, and executing an effective strategy in an effort to match your prediction. In some (analogous) sense, the actual play of the game amounts to exercising cognitive control on your betting cues.

As Rayment (2011) and McLeod (2011) note, the game is scored as follows: If the players on a given team meet or exceed their bet, they receive that bet times ten, plus the amount they went over. Excess hands are counted as 'bags,' which don't have any negative consequence until 10 such bags are accumulated (over the course of several hands, presumably), at which point the team is docked 100 points. In the event, however, that the team does not meet their combined bet, they lose ten times the amount of tricks that they bet. This creates a situation where going under is harmful in the short term while going over may be harmful in the long term, which calls to mind certain temporal biases in utility valuation. Players can also bet 'nil,' or zero, in which case they receive or lose 100 points depending on their success.

The game of Spades has a rich set available cues for betting, but, as in many domains, expert players don't always know exactly how they make their estimate. The goal of the current study was to examine cues involved player betting, which we predicted would rely on things like the utility of a hand and the projected performance of teammates and opponents. This particular domain is interesting because it involves self-prediction. It's very possible, for example, to make a bet that is lower than usual given the apparent utility of a hand, and intentionally play the cards in such a way (namely by throwing away high cards) that you reach the lower bet. In some sense it is possible that experts bet based on their perceived degree of achievement control in a given game.

METHOD

Participants. Four Rice University seniors (3 female) volunteered to participated, without compensation, in the study. The first two players, the *experts*, were very experienced Spades players; playing as a team at least once per week. The remaining two players, the *novices*, had never played Spades before. Additionally, all four subjects indicated having normal or corrected to normal vision.

Design. The dependent variable of interest, or judgement, was the number of hands bid by each player prior to playing each round. While, as mentioned above, there were a large number of cues available to players, the following seven cues were chosen to be included in the lens model based on: (1) observations made during the experiment (i.e. teammate's bid), (2) general importance to the game (i.e., high/low spades/non-spades), and (3) hypothesized importance to players (i.e., cumulative score and previous bid differential).

Cumulative Score. The cumulative score reflected the score each team had at the conclusion of the previous round. The cumulative score for each team was announced at the conclusion of each round, but before bidding began for the next round.

Teammate's Bid. As teams were pairs of players and each player bid individually, the number of hands bid by each player's teammate was recorded for each player.

Previous Bid Differential. At the conclusion of each round, each team had either won the number of hands bid, over bid (and came up short), or underbid (and went over). The previous bid differential reflected the difference between how many team hands were bid and how many team hands were won on the previous round.

High/Low Spades. The number of high (10 through Ace) or low (2 through 9) spade cards each player was dealt prior to bidding each round.

High/Low Non-Spades. The number of high (10 through Ace) or low (2 through 9) non-spade cards each player was dealt prior to bidding each round.

Procedure. At the start of the experiment, two practice games were played in order to orient the novice players to the rules and general flow of the game. The practice games also allowed the experimenters to familiarize themselves with the data collection process. Upon completion of the practice games, a single game of spades was played in which all the relevant data was recorded from each player. Players from each team sat across from one another and each experimenter was assigned to a player. Before bidding commenced each round, the experimenter recorded their player's bet as well as the number of cards, and high/low designations, of each suit dealt to their assigned player. Scoring was continuously updated by the experimenters and read allowed at the conclusion of each round. At the conclusion of the game, each player was given a short demographics questionnaire complete and thanked for volunteering their time.

RESULTS

Novice 1. The participant represented as Novice 1 in the model stated that she used the number of high cards she had to predict how many rounds she would win. As can be seen in Figure 1, Novice 1 did reliably use both the number of high spades and the number of high non-spades in her hand to make her bet. Surprisingly, while the standardized beta coefficient for high spades is strong and positive (.95) as we would expect, the coefficient for high non-spades is negative (-.42). This means that as the number of high non-spades in a hand increases, the bet tends to decrease, which is in contrast to what the participant reported after the experiment was over. The model also shows that beyond just the number of high cards in a hand, Novice 1 uses two other cues reliably to make her bets: previous bid differential and number of low spades. For previous bid differential, the strength is -.62 which means that the further off her team was on the previous hand with their bets, the lower her bet on the next hand.

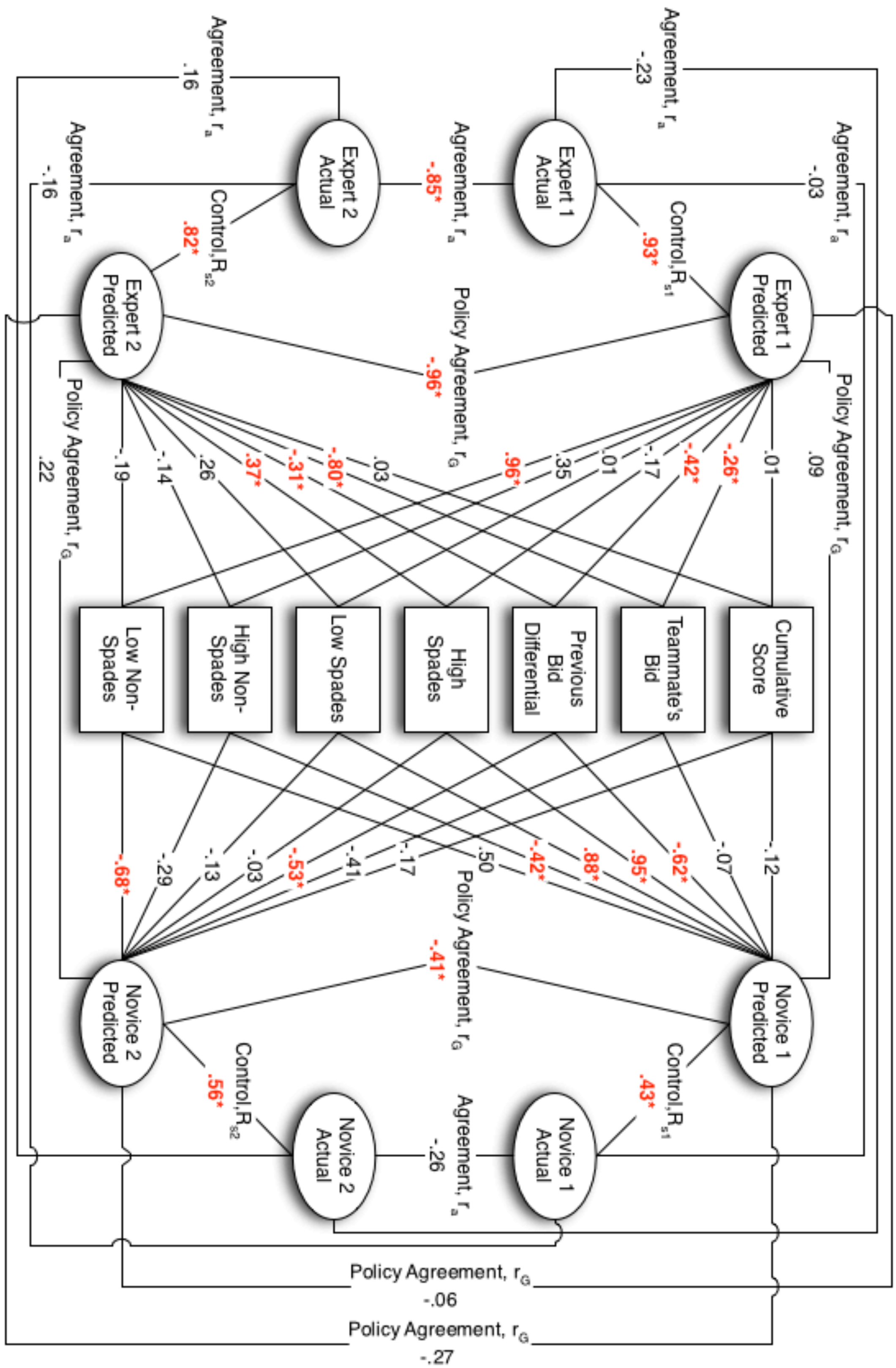

Figure 1. Lens model of expert and novice Spades players. Asterisks denote $p < .05$.

The cue of number of low spades was strongly, positively related (.88) meaning that the higher the number of low spades, the higher the prediction of rounds won. The model also shows that Novice 1 does not use 3 of the cues we studied: the cumulative score (-.12), teammate's bid (-.07), or the number of low non-spades (.50). The other measure that can be taken from just the model of Novice 1 is cognitive control, which is low in this case. The model only accounts for 18 percent of the variance of her bets (.432). This is extremely low and could indicate either that she has very little ability to use the cues to make her judgments or that she is using other cues from the ones we focused on. The latter explanation is weakened by the fact that when we asked Novice 1 what she was using, she did not report anything that we didn't study. Therefore, it is likely that this low cognitive control simply comes from her inconsistency in making bets with respect to the available cues.

Novice 2. Novice 2 reported a similar strategy to Novice 1. When asked how she made her bets, she reported that she looked at the amount of high cards in general, as well as the number of spades in her hand. However, the model shows that only two cues were reliably used by this participant when she made her bets and neither of them were cues about high cards or spades. She did use previous bid differential, and like Novice 1 this cue had a moderately strong, negative coefficient (-.53). Like Novice 1, this indicates that as the discrepancy between the previous bet and previous outcome increased, the bet in the next hand tended to decrease. She also used the number of low non-spades and this cue was also moderately strong and negative (-.68), meaning that the more low non-spades in a hand, the lower her bet tended to be. She did not appear to be using any of the other five cues that we studied: cumulative score (-.17), teammate's bid (-.41), number of high spades (-.03), number of low spades (-.13), or number of high non-spades (-.29). It is very surprising that according to the model, she did not use any of the cues that she claimed to be using when asked after the experiment (high spades, low spades, or high non-spades). However, like Novice 1, Novice 2 showed very low cognitive control so her inconsistency might help to explain the discrepancy between her reported strategy and what the model shows. Although better than Novice 1, the model of Novice 2 only accounts for 31 percent of the variance (.562) of the actual bets, which is still very low. Again this could be due to the use of other cues from what we studied; however the self-report of strategy makes this unlikely. It seems that both of the novices in this experiment show low cognitive control due to inconsistency in making their bets with the information given.

Expert 1. Expert 1 gave a slightly more complicated report of his strategy. He, like the novices, stated that his bet was primarily based on the number of spades and the number of high cards of other suits in his hand. He also reported that if he had a dominant suit in his hand, that this played a role in his betting. However, the model showed that he used 3 cues, and none of them were about the number of spades or high non-spades. Most strongly, he used the cue of low non-spades, which had a strong positive coefficient (.96). This is interesting as it indicates that the higher amount of low non-spades in his hand, the higher his bet. This goes against what he reported as his strategy and also against the logic of the game. The model also shows that he uses his teammate's bid (-.26) and the previous bid differential (-.42). Both of these are negatively related to the bid, so as the teammate's bid and the discrepancy between bet and results from the previous round increase, his bet decreases. The model also shows that he does not reliably use the other 4 cues: cumulative score (.01), number of high spades (-.17), number of low spades (.01), and number of high non-spades (.35). Unlike both of the novices, Expert 1 shows great cognitive control. The model accounts 86 percent of the variance (.932) of his bets, which indicates consistent use of the cues mentioned above. Some of the rest of the variance could come from the fact that we did not have a cue for dominance of a suit, which Expert 1 reported using to make his bets as well.

Expert 2. Expert 2 also reported a more complicated strategy than the novices, stating that she used the presence of kings and aces, the lack of a suit in her hand (defined as 3 or less cards of a particular suit) and having multiple and/or high spades. The model showed three significant cues used by this judge, only one of which was in relation to the cards in a hand. Most strongly she used teammate's bid (-.80). This strong negative coefficient means that as her teammate's bid increased, her bid decreased. She also used the number of high spades (.37), with the greater number of these cards tending to lead to higher bet. Finally, she used previous bid differential (-.31) which, like described previously, means that as the discrepancy between bet and results in the previous round increased, her current bet decreased. She did not appear to use cumulative score (.03), number of low spades (.26), number of high non-spades (-.14), or number of low non-spades (-.19). Like Expert 1, Expert 2 exhibits higher cognitive control than the novices. The model of Expert 2 accounts for 67 percent of the variance of her bets. Again, the rest of the variance might not come from inconsistency like we believe for the novices, but from the fact that we didn't look at some of the cues that she used, such as the lack of a suit in hand.

DISCUSSION

Our hypothesis was that experts would be better at predicting how many hands they would win (and at executing in such a way that supports this prediction), and that such expert bets would be based both on the literal utility of the cards and on other cues like the previous bets, whereas the novices would base their less accurate decisions almost exclusively on a less accurate model of card utility. In addition, we hypothesized that novices would anchor strongly on the modal bet, which is 3 tricks, while experts would be more prone to making novel bets.

Overall both novices exhibited fairly poor cognitive control, most likely due to lack of consistent cue utilization. While it would appear that Novice 1 made her bets primarily based on the number of high cards and that Novice 2 made her bets based off of the number of lows cards, since the overall model predicts the novice behavior fairly poorly, it is likely that these two strategies were inconsistently applied. This is in alignment with expectations about strategic learning in that both novices utilized their previous bid differentials, and may have changed their utility ratings as a result. The explanation is that novices assume that a loss is due to a faulty card utility function, as opposed to a bad result that occurred in spite of a good betting strategy. It may be a case of the representativeness heuristic and failure to assume regression to the mean that produces this type of behavior. However, this sort of assumption may actually be beneficial early in skill acquisition, as there is literally less information available, and novices (who are employing unrefined strategies) are more likely to be performing badly due to a strategic flaw than due to a given hand's result being anomalously poor. The experts both utilized previous bid differential to a much smaller extent, indicating that they were not making drastic strategic modifications. The fact that Novice 2's self report did not agree with her apparent strategy is further evidence of a lack of consistent cue utilization.

The behavior of the experts was more complex overall, for several reasons. Both experts utilized their teammate's bid- something that the novices did not do. This supports our hypothesis that one of the main differences between expert and novice spades playing is due to non-card cues. There are implications here for other scenarios involving cooperative group play. Namely, it is that players may eventually settle on a personal strategy (which might translate into something like dexterity and situational awareness in the domain of sports, or subject knowledge in a focus group), and then attempt to develop models of their teammates. While the experts in our study did not appear to utilize all of the information available to them, they both exhibited high degrees of cognitive control. In addition, informally it seems likely that experts were relying on additional cues like visual signals and the lack of a certain suit. Attendance to the former cue could lead to an improved understanding of teammate intentions, and increased reason to base one's bet on theirs. Attendance to the latter (or other card-utility based cues) could be one of the final pieces of mastery of basic individual self-predictive strategy. In order to test whether this stage-based task learning is feasible, it would be interesting to test the same group of novices longitudinally as they learned the game, or to test alternate groups in between the experts and the novices. However, one caveat to this theory is the fact that the experts in this study had played together before. As such, another followup would be to pair unfamiliar experts, and see if they still relied on each other's bids.

What is particularly fascinating about the game of spades is the sense in which there is no optimal ecologically valid choice. It is a game of self-prediction, but it can also be a game of reading the intentions of one's partner- definitely through the direct measure of their bet and perhaps through more ephemeral signs like reaction, signaling and past experience. Of further note is the way in which our experts differed in their card value cue utilization. Expert 1, for all intents and purposes, appeared to bet irrationally in that he utilized the amount of low cards as a strongly positive cue. However, expert 2 consistently responded to expert 1's high bets, and weighed this much more strongly than any hand-based cue. If one had to characterize this behavior, it would be that both experts had confidence in expert 1's ability to execute a hand in accordance with a given bet, almost irrespective of the cards he had to work with. What this implies is that perceived execution ability (of both oneself and one's teammate) may be an important modulating factor for all other cue weights. In addition, it is possible that expert 1 had strong confidence in expert 2's execution ability in the sense that he thought he could rely on her to underperform based on her cards, thus allowing him to over-perform.

What is very evident from the model is that the expert players play better as a team. Neither of the novice players utilized their teammate’s bid to modify their own. On the other hand, the expert players made strong use of their teammate’s bid. At the risk of reading too much into a small data set, what is evidenced here is that our novices tended to weigh their previous personal performance strongly, and adjust their cue evaluation strategy for this (as evidence by the low cognitive control). Experts, on the other hand, generally have their card evaluation strategy down, and are modulate their bets primarily on the intent and skill of their partner. This indicates that there is actually quite a bit of importance in the actual gameplay phase of spades, and that it is the ability to predict a partner's skill in this that is the primary dividing factor between novices and experts.

REFERENCES

Brunswik, E. (1952). The conceptual framework of psychology. Chicago: University of Chicago Press.

Hammond, K.R., Stewart, T.R., Brehmer, B., & Steinmann, D.O. (1975). Social judgment theory. In M. Kaplan & S. Schwartz (Eds.), Human judgment and decision processes (pp. 271–312).New York: Academic Press.

McLeod, J. (2011, December). *Card Games*. Retrieved from http://www.Pagat.com

Rayment, W. J. (2011, December). *History of the Game of Spades*. Retrieved from http://www.rulesofspades.com/history.php