On TRACS: Dealing with a Deck of Double-sided Cards

Kevin Burns
The MITRE Corporation
202 Burlington Road
Bedford, MA 01730-1420 USA
kburns@mitre.org

Abstract- TRACS (Tool for Research on Adaptive Cognitive Strategies) is a new suite of card games played with a special deck, where the back of each card is a clue to the front of the card. This design simulates the clue/truth structure of real world domains like medicine and warfare, where truths (fronts of cards) must be diagnosed from clues (backs of cards) in order to make decisions (cards to choose, chips to bet, etc.). Here I present the cards and rules of TRACS. I also discuss how the games have been used for computational investigations of memory limits and Bayesian inference. The methods for these studies include human experiments and agent simulations, both of which are facilitated by the unique features of TRACS. The products of TRACS research include a computational model of memory limits and a decision support system for Bayesian inference.

1 Introduction

How do people make decisions and how can systems help them make better decisions in the context of real world domains like medicine, business and warfare? This is the question that motivates my research on computational intelligence, with a focus on cognitive strategies.

Previous research on *Judgment and Decision Making* (JDM, see Connolly et al. 2000) has typically pursued computational studies in the lab. Conversely, previous research on *Naturalistic Decision Making* (NDM, see Zsambok and Klein 1997) has typically pursued ecological studies in the field. While each camp strives for both *rigor* and *relevance*, the reality is that JDM has produced mostly rigorous findings that are lacking in ecological relevance while NDM has produced mostly relevant findings that are lacking in computational rigor.

I think that both rigor and relevance are needed for useful research and that *mind games* can help bridge the gap between JDM and NDM. However, to fulfill this promise, the mind games must be both tractable to computational analysis in the lab (for rigor) and prototypical of psychological challenges in the world (for relevance).

This paper presents a new suite of card games called **TRACS**: **T**ool for **R**esearch on **A**daptive **C**ognitive **S**trategies (Burns 2001, Burns 2005). TRACS is unique in that it uses a special deck of double-sided cards to play games that are, compared to standard card games, more tractable to computational analysis and more typical of practical situations. In this paper, I present the TRACS cards and rules and discuss how TRACS games have been used for basic research on mental models and applied research on support systems.

1.1 Card Games

Card games have two features that make them attractive as mind games for research on cognitive strategies, namely they are *flexible* (for experiments) and they are *familiar* (for participants). Flexibility is desirable so that game tasks can be modified as a research program evolves. Card games are especially valuable in this regard since is it easy to design variations on a theme, as evidenced by new games that have been developed for both recreation and research purposes (Gardner 2001, Abbott 1963). Familiarity to a general population is desirable because a domain-specific mind game (e.g., military or business or medical) requires recruiting and/or training of experts with domain-specific knowledge. Moreover, domain-specific research findings often do not generalize outside the domain. The tokens and rules of card games are abstract analogs of many domains (McDonald 1950), and this allows research findings to be applied across domains.

Besides being flexible and familiar, card games also offer advantages of *rigor* and *relevance*. With respect to *rigor*, the cards in a standard deck present a relatively small set of features (e.g., suits and pips) to constrain the possibilities that players must consider as they reason about probabilistic game states in a dynamic context. With respect to *relevance*, card games are played with imperfect information about face down cards, which simulates real world conditions better than board games like checkers and chess that are played with perfect information about the game state. With respect to *both* rigor and relevance, card games are better than 3-D graphic and virtual adventure games for research on human judgment and decision making – because the simple images (cards) and discrete sequences (moves) allow players and researchers to focus more directly on cognitive strategies rather than sensing and motor skills.

1.2 New Games

Along with the advantages noted above, card games also have some disadvantages. With respect to computational *rigor*, the optimal decisions in most card games are not amenable to analytical solution (Epstein 1977, Koller and Pfeffer 1997), except for trivial versions (Kuhn 1950, Nash and Shapely 1950) that are not very *relevant* to practical situations. The combinatorial complexities of most cards games played with a standard deck of 52-cards make it extremely difficult to establish normative performance, which is needed to benchmark cognitive strategies. And, with respect to *relevance*, standard playing cards provide information on only one side of the cards, which does not reflect the basic clue/truth structure of many real world problems like diagnosing a medical disease or military target (truth) from an X-ray image or radar return (clue).

The TRACS cards and rules are designed to overcome both of these limitations found in standard card games. For practical *relevance*, TRACS uses double-sided cards where the backs give clues to the fronts, and TRACS poses diagnostic and decision making challenges that simulate real world dilemmas. For computational *rigor*, TRACS is a suite of games ranging from simple (and tractable) to complex (less tractable), which facilitates progressive research on cognitive strategies.

2 The Cards

Standard playing cards present information on only one side, in the form of shape-color *suits* (Club, Diamond, Heart, Spade) and numerical *pips* (A, 2, 3, …, J, Q, K). The TRACS cards (Figure 1) are different because they present information on both sides to better reflect the clue/truth structure of imperfect information in real world domains. This novel feature of TRACS provides research advantages over standard card games, but also presents a design challenge in helping players to internalize the structure of the double-sided (unfamiliar) deck.

The design of the deck is based on the notion of *tracks* and *treads*, where the back of each card is a track that gives a clue to the tread on the front. The analogy is that of a track (shape) left by the tread of a shoe or a tire. Each tread is set of shapes and there are two treads in TRACS: a Red tread (Figure 1, upper) and a Blue tread (Figure 1, lower).

Each track, on the back of a card, is a single black shape (triangle, circle or square) that gives a clue to the tread on the front of the card. The clue comes from the structure of the deck, in which there are different numbers of each track (shape) in each tread (set), as shown in Figure 1 and as illustrated by the Red and Blue sets of shapes in the center on the front of the cards.

2.1 Rationale

Why are there only two treads (Red and Blue)? Two treads are used to keep the games as simple possible but still interesting. If there were only one tread the player would have nothing to diagnose.

Why are there three tracks? Again, it is to keep the games as simple as possible but still interesting. Three is the minimum number of track types needed to capture the basic difference between what is *likely*, *unlikely* and *ambiguous* (50-50) in probabilistic diagnoses.

How is this structure relevant to real world domains? A fundamental dilemma in virtually every domain is to diagnose the likely truth from a given clue. For example, in medical diagnosis one must infer the most likely state of a tissue (healthy or diseased?) from the clue given by an X-ray image. Similarly, in military intelligence one must infer the most likely identity of a possible target (friendly or enemy?) from the clue given by a radar return. TRACS *cards* reflect the essential structure of this task, because players must infer the likely truths (treads) from clues (tracks). TRACS *rules* reflect the relevant context in which people must deal with clues and truths in the real world, which is both probabilistic and dynamic.

Along with this basic clue/truth structure, a related feature of the TRACS deck is that it contains multiple copies of each track/tread card type, i.e., either 2, 4 or 6 of each card type (see Figure 1). This design has both theoretical advantages and practical advantages over standard playing cards. A theoretical advantage is that TRACS can be used to study the memory processes that people use to count “carbon copies” (multiple instances), as opposed to unique objects like the cards in a standard deck. A practical advantage is that the deck can be scaled (halved, doubled, sampled, etc.) to change the number and/or distribution of card types. For example, the deck used in solitaire games can be doubled for a two-player game to preserve the number of cards per player.

2.2 Example

As an example of the basic problem that arises in TRACS (and the real world), consider a card that is dealt from a full deck with tread down and track up. Assume this card shows a triangle track.

Based on the distribution of track/tread cards in a full deck (Figure 1), the Red:Blue odds for this triangle are 6:2 [P(Red)=6/8=75%]. Similarly, the Red:Blue odds for a square track are 2:6 [P(Red)=2/8=25%]. Thus, when faced with a decision like “choose the Red card”, you would do better to pick a triangle track than a square track.

Now consider a situation where some triangle tracks have been turned over in play but then removed from play and taken out of sight. How well can you remember which cards you have seen so you can update the track/tread odds for a triangle track that is dealt later in the game?

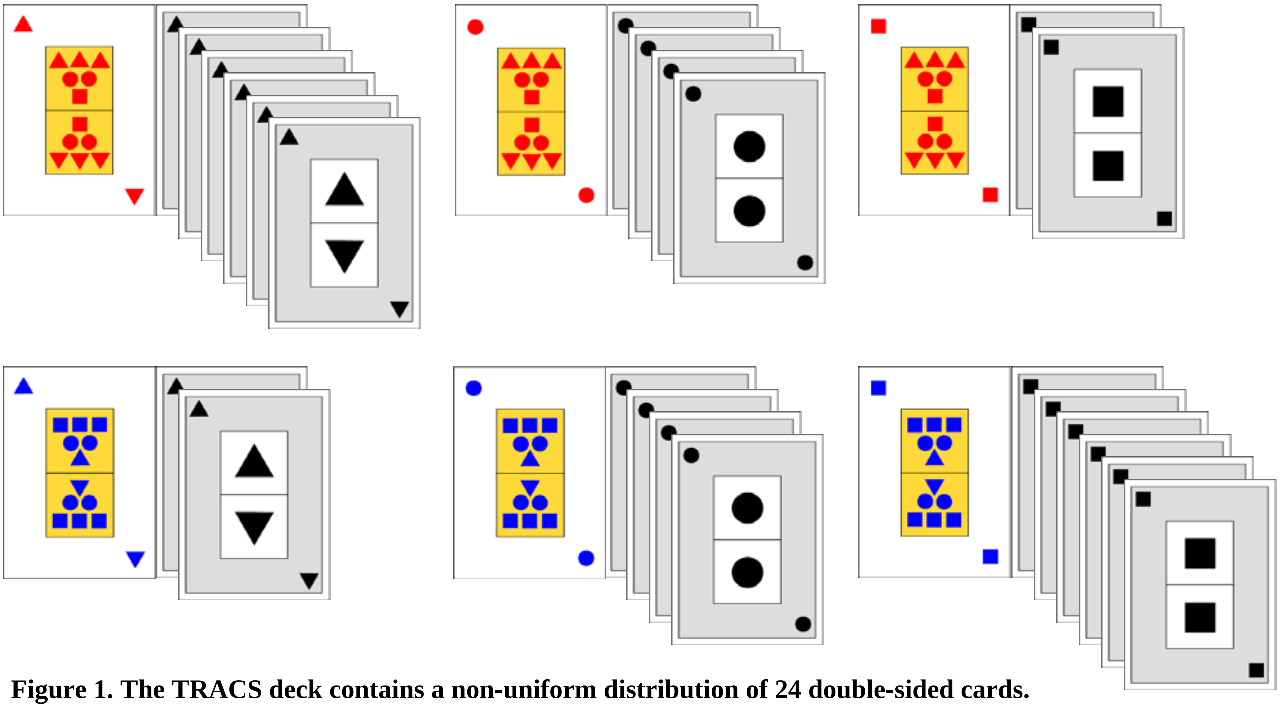


Figure 1. The TRACS deck contains a non-uniform distribution of 24 double-sided cards.

Next consider a situation where you get additional information from a *spy* who, like all spies, is of limited reliability. For example, assume a triangle track is dealt at the start of the game and the spy says, “That track is Blue”. Also assume that the reliability of the spy is known (e.g., 90% correct in reporting Red or Blue). How do you use this *likelihood* information along with your *prior* knowledge of the Red and Blue probabilities (based on the deck distribution) to estimate the *posterior* (after spy) probability that the triangle track is Red?

Finally, consider a situation where a triangle track appears in the hand of an opponent, along with two square tracks. Also assume that the game is a 3-card poker where the best hands are flushes (all three Red cards or all three Blue cards) and that your opponent, who holds the triangle and two squares, has just made a big bet. Based on the distribution of cards in the deck, his two squares are probably Blue and his one triangle is probably Red. But he just made a big bet and this behavior provides some evidence to support the hypothesis that his cards are either all Red or all Blue. How well can you fuse the evidence from his tracks and bets to make inferences about what he has in his hand, so you can make the best choice about whether to call his bet, raise his bet or fold your hand?

In fact, these three examples (above) highlight the cognitive challenges of three TRACS games, called *Straight TRACS*, *Spy TRACS* and *Poker TRACS*, respectively. In the following sections I discuss these and other TRACS games.

3 The Games

TRACS is a suite of games that can be played with real cards or online (Burns 2001, Burns 2005). The online games are programmed in Java, and computer versions have also been written in MATLAB for agent simulations and human experiments.

The TRACS deck can be used to play many games, both solitaire and multiplayer, only some of which have been programmed in computer versions. One class of online games, the ***TRACS Arcade,*** includes three solitaire games called *Straight TRACS*, *Wild TRACS* and *Booty TRACS*. *Straight TRACS* is a game of forced choice on each turn, played in a series of turns. *Wild TRACS* gives the player some options on each turn, and *Booty TRACS* gives the player even more options on each turn.

The other class of online games, in the ***TRACS Casino***, includes gambling games for one or more players. These games are similar to familiar casino games but with various twists that are made possible by the unique design of TRACS. The games include *Slot TRACS* (like a slot machine), *Black TRACS* (like Blackjack) and *Poker TRACS*.

Below I discuss the games in the *TRACS Arcade* with a focus on *Straight TRACS*, which is the simplest game and which has been used to perform normative analyses and cognitive experiments. I also discuss the games in the *TRACS Casino* with a focus on *Poker TRACS*, which is the most complex game and is therefore the most computationally interesting. The details of all games, as well as online play, are available at www.tracsgame.com and http://mentalmodels.mitre.org.

4 TRACS Arcade

The *TRACS Arcade* is a suite of solitaire games. The simplest game, called ***Straight TRACS***, is a matching game played in a series of turns as the deck is depleted.

Figure 2 illustrates a typical turn. At the start of the turn, three cards are dealt from the deck (on left) to a *field* (on right) as one *tread* flanked by two *tracks* (Figure 2a). The diagnostic challenge is to judge the likely color of each track. The decision challenge is to choose the track, either left or right, that is most likely to match the color of the tread in the middle. The chosen track is turned (Figure 2b) and the turn is scored as a *save* (if colors match) or a *strike* (if colors mismatch, as in Figure 2b). The pair of treads (save or strike) is then removed from play and stored in a stack (tread down, not shown in Figure 2), using one stack for saves and one for strikes.

The remaining track on the field is turned to reveal its tread (Figure 2c) and this tread is moved to the center of the field where it becomes the color to match on the next turn (Figure 2d). Two more tracks are dealt from the deck to the field (left and right) and the sequence continues until all cards have been paired as either a match (save) or mismatch (strike). The last pair does not count because the player has no choice.

The object of the game is to make saves (matches) and avoid strikes (mismatches). The challenge of the game is to count the cards and update odds as the deck is depleted, in order to make the best choice on each turn.

Straight TRACS is a good game for learning TRACS, but it is not too fun for participants and not too deep for experiments. This is because every turn is a forced choice (left or right) with a fixed outcome (strike or save), whereas choices in real life often involve several options and scalable outcomes. Thus, the *TRACS Arcade* also includes two variants of *Straight TRACS* that are designed to capture these features of real world decisions.

In one game, called ***Wild TRACS***, the player gets four wild cards that can be used as tickets to exercise options. Each wild card can be played in one of two ways, either as a *spare* (which is like passing a turn) or a *dare* (which is like betting double-or-nothing). Each wild card can be played only once, as either a spare or a dare, and the player's challenge is to optimize use of this limited resource and finish a trip through the deck with no strikes. In another game, called ***Booty TRACS***, every *save* (match) becomes *booty* that the player can use to bet double-or-nothing on later turns. In this way, a player must bet booty to get booty, and the player can scale the stakes (booty) beyond the one point of a save or strike in *Straight TRACS* and the additional point of a dare in *Wild TRACS*. The object is to get the most booty on a trip through the deck.

The simplest game of Straight TRACS has been used to perform human experiments and agent simulations, and the findings are discussed in Section 6. Section 5 (below) outlines other games in the TRACS Casino.

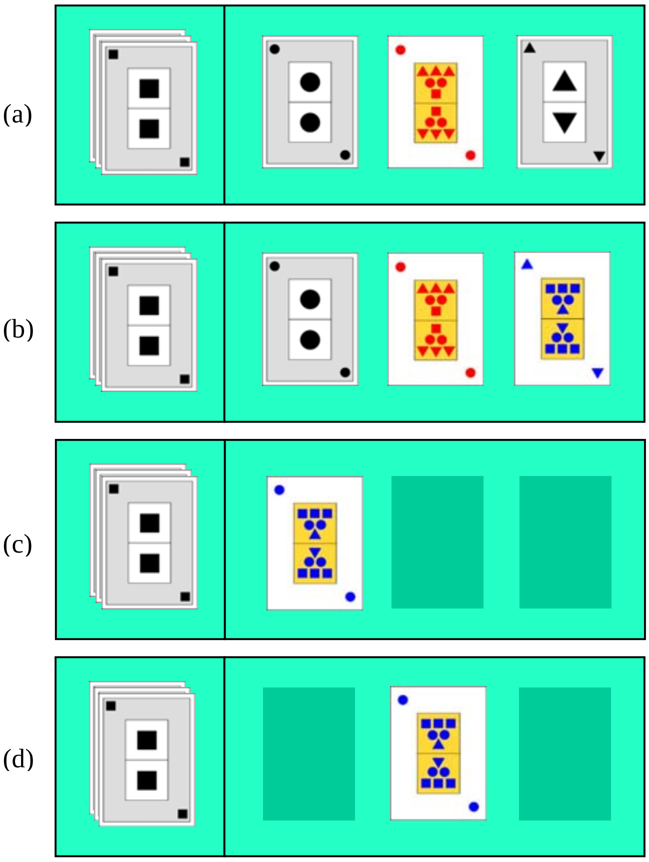
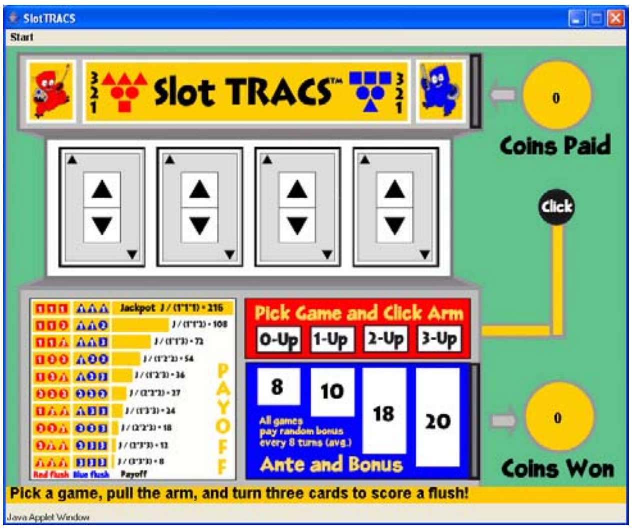


Figure 2. A turn in *Straight TRACS*. This turn is scored a *strike* (colors mismatch).



All games pay random bonus every 8 turns (avg.)

Pick a game, pull the arm, and turn three cards to score a flush!

Figure 3. Screen shot of *Slot TRACS*.

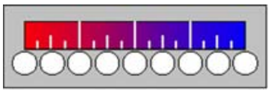


Figure 4. Colored ruler used as a probe in human testing on *Straight TRACS*.

5 TRACS Casino

The *TRACS Casino* is a suite of gambling games, similar to familiar games but with a TRACS twist. ***Slot TRACS*** is the simplest game, and is similar to slot machines except the player must make a few decisions other than whether or not he wants to pull the arm, which is the only decision in most slot machines. *Slot TRACS* (Figure 3) has four windows (slots), and behind each is a simulated wheel with a full TRACS deck (see Figure 1). The wheels are spun and when they stop a single card appears behind each window. The player must select three cards in three of the four windows, trying to make a flush of all three Red or all three Blue. The selected cards are clicked to turn them face up and show their colors. The player gets a payoff if he makes a flush, and the payoff gets higher as the flush gets rarer, where rareness is determined by the distribution of cards in the deck. For example, a flush of three Red squares or three Blue triangles is the most rare and so it pays the Jackpot. Other flushes pay less in proportion to their rareness.

Slot TRACS is interesting because the player must make two kinds of decisions. First, he must decide how many cards he wants to be spun face up: 0, 1, 2 or 3. The ante (cost to play) increases as more cards are spun face up and the player must balance the cost of this extra information with its benefit in making rare flushes.

The second decision in playing *Slot TRACS* is which track(s) to turn face up after the spin. This choice is interesting from a cognitive perspective because it involves a tradeoff between a high chance of getting a low payoff versus a low chance of getting a high payoff. The payoff schedule in *Slot TRACS* is purposely designed such that all choices are normatively equivalent, yet players tend to exhibit individual preferences that reflect risk seeking or risk averse tendencies.

Black TRACS is another game in the *TRACS Casino*, and it is similar to Blackjack but with a twist. Like Blackjack, *Black TRACS* is played against a dealer who must make his choices according to a pre-set strategy while the player can adjust his strategy to account for changes in the number of each card type remaining in the deck. Unlike Blackjack, it is harder to keep an effective count in *Black TRACS* because the player must count six types of track/tread cards at once. In the standard card-counting strategy for Blackjack (Lewis 2002), the player need only count only one thing, namely the net number of high cards that have been seen (i.e., high cards seen minus low cards seen).

Poker TRACS is the most complex and challenging game in the *TRACS Casino*. Unlike standard poker, a hand is made of three cards (not five), and the ranking system is the same as the payoff structure of *Slot TRACS* (see above). That is, a flush (three cards of the same color) ranks higher than a non-flush, and rarer flushes are better, where rareness is determined by the number of each track/tread card type in the full deck.

More formally, the possible flushes are ranked by their conditional probability P(flush|tracks) in the full deck, and the ranks are computed as follows: First, each card is assigned a number 1, 2 or 3 proportional to the number of its track (shape) in the tread (set) as illustrated by the two tread designs (Red and Blue) on the fronts of the cards. That is: Red triangle = 3, Red circle = 2, Red-square = 1; Blue square = 3, Blue circle = 2, Blue triangle = 1. Then, for a flush of 3 cards {X, Y, Z}, all Red or Blue, the numbers are multiplied to get N=X*Y*Z. The 10 possible N are {1, 2, 3, 4, 6, 8, 9, 12, 18, or 27}, where smaller N are rarer. [N=1 is possible only with a double deck.]

This ranking system preserves the basic structure of standard poker but eliminates the need to memorize a ranking schedule (like Flush beats Straight) and reduces the number of possible hand strengths by orders of magnitude. That is, while standard poker also has 10 ranks based on rareness (i.e., Royal Flush, Straight Flush, Four-of-a-Kind, Full House, Flush, Straight, Three-of-a-Kind, Two Pair, One Pair, No Pair), it makes further distinctions within each rank (e.g., King beats Queen even though each is equally rare). *Poker TRACS* reduces the number of possibilities to 10 flushes (plus non-flushes), and the flush ranks makes it easy for players to compute the odds of hands that their opponent may hold.

For example, in standard poker (with 5-card hands dealt from a 52-card deck), the odds that your opponent was dealt a Straight versus a Flush are about 2:1 (Straight:Flush). But this is a complex calculation that poker players cannot do in their heads, and many players do not even know that the approximate answer is 2:1. In *Poker TRACS*, you can easily estimate the odds of possible hands that your opponent was dealt because these odds are given by the tracks in his hand and the ranking system (above). For example, if your opponent's tracks are {triangle, square, square}, then either he holds a Red flush of N=3*1*1=3, or a Blue flush of N=1*3*3=9, or a non-flush. Furthermore, it is approximately three times less likely (3:9) that he holds the Red flush with N=3 than the Blue flush with N=9.

I say "approximately" because the ranking structure reflects the baseline (full deck) odds and the actual odds of flushes in opponents' hands change for two reasons. First, the fronts of some cards may be observed in play (i.e., in your own hand). For example, if you hold two Red squares, and there are only two in the whole deck, then any squares in your opponent's hand must be Blue. Second, you get additional information about your opponent's hand from his bets and raises. In the above example, if your opponent makes a big bet or raise then he probably has the stronger flush rather than the weaker flush or a non-flush, unless he is bluffing. And, of course, this is the essence of poker (and business and warfare, see McDonald 1950), in which players must use imperfect information from cards and bets along with models of their adversaries to make inferences about one's chances of winning and to make investments of one's chips to the pot.

6 Human Testing

Agent-based simulations have been performed in order to establish normative strategies in *Straight TRACS* and other solitaire games, and in order to develop cognitive surrogates (robot players) for multiplayer games like *Poker TRACS*.

Human-based experiments have been performed on the simplest game of *Straight TRACS* and a variant called *Spy TRACS* in order to measure cognitive performance against normative standards. The methods and findings from these experiments are reviewed below, to show how TRACS can be used to develop valuable insight into how well people reason about probabilistic information in dynamic situations.

6.1 Experiment 1

Experiment 1 was designed for two purposes. One purpose was to test how well people could count cards and update odds in *Straight TRACS*. Another purpose was to develop a computer model of cognitive limits.

Straight TRACS requires probabilistic estimation of track/tread odds under dynamic conditions as the deck is depleted, where the key skills are basic memory (like card counting in Blackjack) and revising probabilities. The hard part is that, in TRACS, the player must keep a count of all six track/tread cards types (see Figure 1) in order to update the odds. For example, the Red:Blue odds for circle tracks start out at 4:4, and to update the odds after some circles have been seen the player must know both how many Red circles have been seen and how many Blue circles have been seen.

The probe for Experiment 1 was a colored ruler with buttons (Figure 4), which appeared on the computer screen below each track. The ruler ranges from 100% Red on one end, to 50-50 in the middle, to 100% Blue on the other end. The player's task was to click the button that matched the chances (player judgment) that the track will turn out Red or Blue. The experiment also measured which track the player chose to match the color of the tread in the middle (see Figure 2). Here I focus on the data for players' judgments (of % Red or Blue) rather than on players' choices (of track to turn). Further details on both judgments and choices in the experiment are provided elsewhere (Burns 2002, 2003).

The major finding from Experiment 1 is that people are extremely limited in their ability to count cards and update odds. Approximately 100 participants were tested, each playing 10-20 games of *Straight TRACS*, where each game involved 11 turns. All participants played several practice games beforehand, and some participants played many practice games. The cognitive performance of participants was seen to exhibit a pattern of *anchoring and adjustment* (Burns 2002, 2003), in which players were anchored to the baseline (full deck) odds and made only minor adjustments thereto – typically much less than the optimal adjustments – as a result of memory limitations.

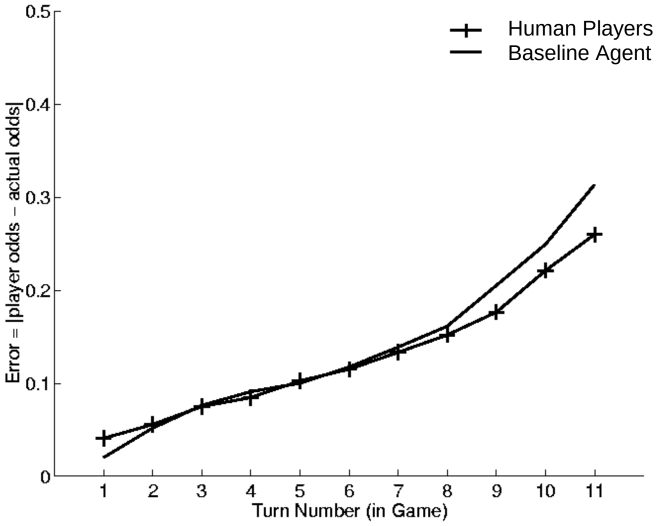


Figure 5. Error versus turn in *Straight TRACS*, showing a baseline bias.

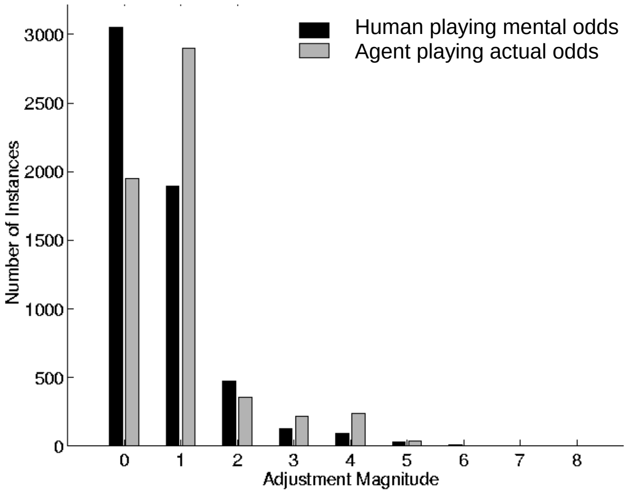


Figure 6. Turn-to-turn adjustments in odds, showing anchoring (adjustment magnitude 0).

Figure 5 shows the average performance for human players and for a simulated baseline agent, who is a player that never updates odds. A perfect player who correctly updates odds at each turn would have zero error always. This figure illustrates the *baseline bias* that people exhibit in playing *Straight TRACS*, i.e., the average error increases as the deck is depleted in play similar to the behavior of a baseline agent.

Figure 6 shows the number of turn-to-turn adjustments in odds made by human players and perfect agents. Here the magnitude of adjustment is measured as a span of buttons on the colored ruler (see Figure 4). Figure 6 illustrates *anchoring* behavior where human players make too many *non-adjustments* (adjustment magnitude zero).

Besides these results, I also measured the hits and misses made by players in counting all cards of a certain class, where the six card types (Figure 1) are grouped into three classes (2-count, 4-count, 6-count) depending on how many of that card type there are in the deck. A hit occurred when all of the Red or all of the Blue cards of a given track type (triangle, circle or square) had been turned over in play and the player correctly clicked the rightmost (Blue) or leftmost (Red) button on the colored ruler. A miss occurred when all of the Red or all of the Blue cards of a given track type had been turned over and the player did not click the rightmost (Blue) or leftmost (Red) button. The results show a hit rate of ~50% for counting up to two of each card type (while counting all six types at once), but the hit rate for counting up to four or six was only ~10%.

6.2 Cognitive Model

Based on these findings, I developed a computational model to explain and predict human memory in playing *Straight TRACS* (Burns 2003). The model employs *fuzzy functions*, which are like fuzzy logic in that they specify a fuzzy mapping from a physical event (turning a card) to a cognitive belief (update of odds). The model is based on an *accumulator analogy* where discrete bins are filled in a stepwise fashion and the filling of one bin after a lower bin is governed by a fuzzy function (leaky filling).

Figure 8 shows how well the computer model compares to the cognitive data for all three track types (triangle, circle, square) in one game. The black plots the mean (line) and standard deviation (bars) of all data from humans. The gray plots the mean and standard deviation for the model, which exhibits variability due to its fuzzy functions. The dotted line plots the performance of a perfect agent who counts cards and updates odds with no error. The comparison shows that the model (gray) does a good job of capturing the mean and spread of cognitive performance (black), especially relative to normative performance (dotted). Results were similar for other games as reported elsewhere (Burns 2003).

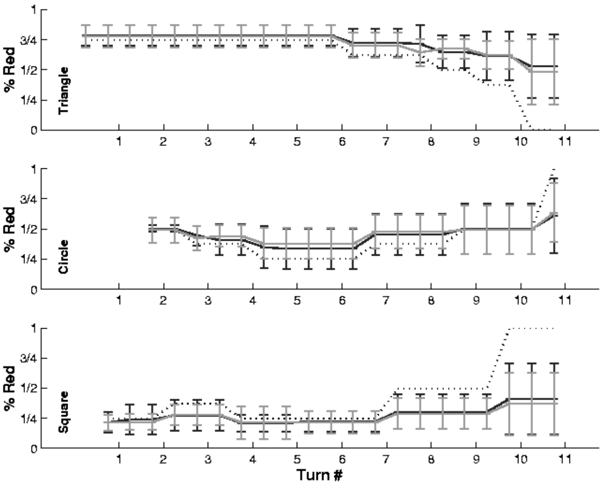


Figure 7. Model (gray) versus data (black).

These experimental findings and the computational model demonstrate how TRACS can be used to gain insight into human thinking. Both descriptive data and predictive models are needed to understand the strengths and limits of cognitive competence, and Experiment 1 shows that TRACS can be used as a test bed for getting such data and building such models.

While memory limitations are important in many practical applications, computer systems that can support such limits are straightforward (e.g., a card counter in *Straight TRACS*). Thus, further experiments (discussed below) were designed and performed to study another task where cognitive competence is also limited but where a support system is not so obvious.

Experiment 2 used a game called *Spy TRACS* to study *Bayesian inference*, which is a critical and challenging task that arises in many real world domains. Like *Straight TRACS*, the basic problem in *Spy TRACS* is to update odds in light of additional information. But unlike *Straight TRACS*, the additional information in *Spy TRACS* comes from a “teammate” (spy) – who like all teammates is of limited reliability

7 System Design

To test cognitive competence in a task of Bayesian inference, *Spy TRACS* is played like *Straight TRACS* but with two twists. One twist is that the player does not have to count cards because he is given the *deck odds* based on the current card count by the computer. The other twist is that the player is also given a spy report with associated reliability for each track. The player’s task is to combine the deck odds with the spy odds to get the *fused odds* in order to make the best choice of a track (left or right) on each turn. A colored ruler like that of Figure 4, but with numerical % instead of buttons, was used for presenting odds (deck and spy) to players and for measuring odds (fused) from players.

The results (Burns in press) confirmed previous findings of cognitive conservatism (Edwards 1982) in Bayesian inference. That is, people extracted far less certainty than they should have from the data they were given. For example, given a deck probability of 80% Red and a spy probability of 67% Red, most people reported a fused probability of less than 80% Red, and some people reported <67% Red. The Bayesian answer is actually 89% Red.

The findings from *Spy TRACS* were used to build and test a support system that could help people in tasks of Bayesian inference. The support system, called *Bayesian Boxes* (Burns in press, 2004) works like a colored calculator (see Figure 8) where the user dials in a *prior* (deck odds, at bottom) and a *likelihood* (spy odds, on the left) and reads off the *posterior* (fused odds, at the top). Further experiments were performed with *Spy TRACS* to evaluate the benefits of *Bayesian Boxes*.

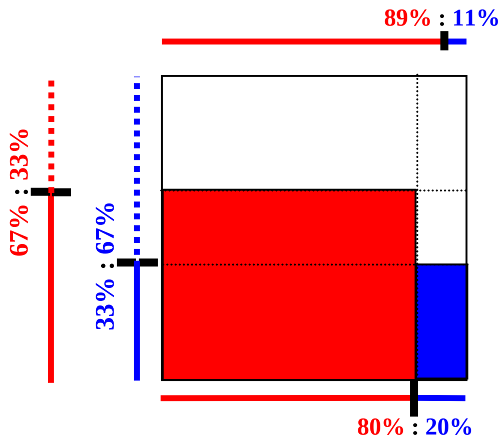


Figure 8. Screen shot from a colored calculator called *Bayesian Boxes*.

Here the intent was not to see if people could dial numbers in and read numbers off the colored calculator. Rather, the intent was to see if practice with *Bayesian Boxes* improved people's intuitive understanding of Bayesian inference. So, the experimental procedure was as follows.

After several games of *Spy TRACS* to get the *before* data (which motivated the development of *Bayesian Boxes*, see above), players were given the colored calculator along with a few sample problems to solve with it (along the lines of the problems they had faced in *Spy TRACS*). After about 5 minutes of practice, the colored calculator was taken away and players were tested on several more games of *Spy TRACS* to collect *after* data. The *before* data was compared to the *after* data and the results showed a significant increase (about doubling) in the fraction of Bayesian responses (Burns in press). Encouraged by these results, *Bayesian Boxes* has been further enhanced and applied to real world problems of Bayesian inference like the forensic assessment of disputed authorship (Burns 2004).

8 Conclusion

TRACS is a suite of games played with a special deck of double-sided cards. In this paper I discussed how TRACS can provide a unique blend of rigor and relevance for research on cognitive strategies. I also reviewed how TRACS has been used to perform cognitive experiments and computer simulations, and how TRACS can be used to develop designs for support systems in practical applications.

Acknowledgments

This research was supported by the MITRE Technology Program. Thanks to the many MITRE volunteers who participated in TRACS experiments.

Bibliography

Abbott, R. (1963) Eleusis. In Abbott, R. *Abbott's New Card Games.* New York, NY: Funk & Wagnals, pp. 73-92.

Burns, K. (in press) Dealing with Probabilities: On Improving Inferences with Bayesian Boxes. In Hoffman, R., (Ed.) *Expertise Out of Context*. Mahwah, NJ: Lawrence Erlbaum.

Burns, K. (2005) *Mental Models in Naturalistic Decision Making,* http://mentalmodels.mitre.org.

Burns, K. (2004) A Bayesian Approach to Detecting a Hoax. *Proceedings of the International Conference on Knowledge Engineering and Decision Support*, Porto, Portugal, July 21-23, pp. 305-312.

Burns, K. (2003) 1, 2, 3, More: An Accumulator Architecture for Anchoring and Adjustment. *Proceedings of the 7th Joint Conference on Information Sciences*, Cary, NC, September 26-30, pp 1673-1676.

Burns, K. (2002) On Straight TRACS: A Baseline Bias from Mental Models. *Proceedings of the 24th Conference of the Cognitive Science Society*, pp. 154-159.

Burns, K. (2001) *TRACS: A Tool for Research on Adaptive Cognitive Strategies*, www.tracsgame.com.

Connolly, T., Arkes, H., and Hammond, K. (2000) *Judgment and Decision Making: An Interdisciplinary Reader*. Cambridge, UK: Cambridge University Press.

Edwards, W. (1982) Conservatism in Human Information Processing. In Kahenman, D., Slovic, P., and Tversky, A. (Eds) *Judgment under Uncertainty: Heuristics and Biases.* New York, NY: Cambridge University Press, pp. 359-369.

Epstein, R. A. (1977) *The Theory of Gambling and Statistical Logic*. San Diego, CA: Academic Press.

Gardner, M. (2001) The New Eleusis. In Gardner, M., *The Colossal Book of Mathematics*. New York, NY: Norton, pp. 504-514.

Koller, D., and Pfeffer, A. (1997). Representations and Solutions for Game-Theoretic Problems. *Artificial Intelligence* 94:167-215.

Kuhn, H. W. (1950) A Simplified Two-Person Poker. *Contributions to the Theory of Games*, Vol. 1. Princeton, NJ: Princeton University Press, pp 97-103.

Lewis, K. (2002) How to Count Cards and Beat Vegas. In Mezrich, B., *Bringing Down the House: The Inside Story of Six MIT Students Who Took Vegas for Millions*. New York, NY: Free Press, pp 252-257.

McDonald, J. (1950) *Strategy in Poker, Business and War*. New York, NY: Norton.

Nash, J. F., and Shapley, L. S. (1950) A Simple Three-Person Poker Game. *Contributions to the Theory of Games*, Vol. 1. Princeton, NJ: Princeton University Press, pp 105-116.

Zsambok, C. E., and Klein, G., (1997) *Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum.