Dealing with TRACS: The Game of Confidence and Consequence

Kevin Burns

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

Abstract

Chance discovery and management require judgments and decisions under uncertainty. Our research explores how, and how well, people make these judgments and decisions. Our tool is a card game called TRACS in which players are faced with prototypical tasks of "probabilistic risk assessment" and "dynamic resource management". Our methods include experiments with human subjects to identify cognitive strategies, and simulations with software agents to evaluate the effectiveness of cognitive and alternative strategies. These scientific investigations on TRACS have potential applications to "command and control" in a variety of practical situations. This paper provides the conceptual foundation, presents our initial investigations and previews practical applications of TRACS.

Introduction

Chance discovery and management are a "game" where people must assess their "confidence" in possible events and forecast the "consequences" of possible options. We are interested in the cognitive challenges of this game from a decision support perspective.

More specifically, we seek to understand the strengths and limits of human thinking in order to identify opportunities for automation and advising. On the one hand, we seek to characterize mental strengths so they can be implemented in autonomous agents. On the other hand, we seek to characterize mental limits so they can be supplemented with advisory systems.

We are investigating human performance in chance discovery and management with TRACS: A Tool for Research on Adaptive Cognitive Strategies (Burns 2001a; www.tracsgame.com). TRACS is a family of card games played with a special deck of two-sided cards. Compared to standard playing cards, the back/front structure of TRACS cards better captures the clue/truth structure of practical domains where people must make diagnoses and decisions (about the fronts of the cards) with only partial information (given the backs of the cards). Fortunately for researchers, the constraints added by the backs of the cards also make TRACS more tractable (than other card games) to mathematical analysis of optimal strategies.

With this dual advantage of practical relevance and mathematical rigor, TRACS provides a useful "micro world" for modeling and measuring human judgment and decision making (Burns 2001b, Burns 2002). Our work on TRACS includes experiments with human subjects to identify cognitive strategies, and simulations with software agents to evaluate the effectiveness of cognitive and alternative strategies. Our goal is to gain insight into how people manage the limited resources of their minds (internal) and their worlds (external) to achieve desired results in uncertain endeavors.

This paper outlines the conceptual foundations of TRACS, to explain how the laboratory game captures fundamental challenges of practical domains. We then present the results of our initial investigations, which measure human performance in a game of "risk assessment". To close, we discuss the potential applications of our work to the design of decision support systems and training.

Foundation

The complex challenges of command and control are ubiquitous in industry (e.g., process management), medicine (e.g., disease management) and defense (e.g., battle management). While the specific details vary from domain to domain, the prototypical tasks of "naturalistic decision making" (Klein 1998) in these and other domains include: (1) Risk Assessment, given uncertain information, (2) Resource Management, over extended missions, and (3) Rational Engagement, in collaboration and competition.

TRACS is designed to capture these prototypical tasks in a recreational game (Burns 2001a, Burns 2001b). The game is played with a deck of 24 two-sided cards (Table 1) that can be adjusted (doubled, sampled, etc.) to accommodate various numbers of players. The backs of the cards show black shapes, called "tracks", and the fronts of the cards show colored sets (of the same shapes), called "treads". The tracks provide probabilistic clues to the treads based on the distribution of cards in the deck. The initial distribution (Table 1) defines the "baseline" odds, but this distribution changes dynamically during the game.

There are a variety of different TRACS games, each designed to address specific research questions. Below we outline the rules for three games that each focus on one task of naturalistic decision making. “Straight TRACS” focuses on Risk Assessment, “Wild TRACS” focuses on Resource Management and “Spy TRACS” focuses on Rational Engagement.

Table 1: Distribution of 24 cards in the TRACS deck.

# of Cards	6	4	2	2	4	6
Front (tread)	**Red**	**Red**	**Red**	**Blue**	**Blue**	**Blue**
Back (track)	▲	●	■	▲	●	■

Straight TRACS: The Basic Game

Straight TRACS can be played alone or with others. To play solitaire, the deck is held face down and three cards are dealt to a field. Two cards are dealt face down, showing their tracks, and the third card is dealt face up, showing its tread. The player turns over one of the two tracks to reveal its tread, trying to match the tread (color) of the third card. The turn is scored a “save” if the treads match or a “strike” if the treads mismatch. The two treads are removed from the field and the remaining track is turned to reveal its tread, which becomes the tread to be matched on the next turn. Two new tracks are dealt from the deck, a track is turned and scored, etc., until all cards (except the last two, which do not count) have been paired.

Straight TRACS is a game of Risk Assessment. The object is to minimize risk (strikes) on a trip through the deck. The challenge is to count cards and update odds in order to choose the best track on each turn. The question is how, and how well, do people update odds and make their choices? Our initial experiments and simulations (see below) were performed using this game.

Wild TRACS: The Options Game

Wild TRACS goes beyond Straight TRACS by giving the player a limited number of “wild cards” that can be used to prevent or mitigate risk (strikes). Each wild card is a token that can be played as either a “spare”, for risk prevention, or a “dare”, for risk mitigation. A spare allows the player to turn over a track without incurring a strike. Regardless of the outcome, match or mismatch, the turn is scored a save. A dare allows the player to bet, double-or-nothing, on the outcome of the current turn. If the turn is a save, then a previous strike is reversed and counted as a save. But if the turn is a strike, then a previous save is reversed and counted as a strike.

Wild TRACS is a game of Resource Management. The object is to finish a trip through the deck with zero strikes. The challenge is to optimize the use of wild cards to prevent and mitigate risk. The question is how, and how well, do people plan and use their limited resources? In particular, how do people manage the tradeoff between risk prevention (spares) and risk mitigation (dares)?

Spy TRACS: The Fusion Game

Spy TRACS is a game of Rational Engagement, played with simulated teammates as spies. Like Straight TRACS, the object of Spy TRACS is to avoid strikes. Like Wild TRACS, the player is given a limited number of resources (spies) to be used during the game. When used, a spy provides the player with a report (% Red or % Blue) on the tread (color) of a track. The player must then combine this “evidence” with his “prior” belief to make a “posterior” judgment about the best track to turn. The object is to allocate spy resources and interpret spy reports in order to minimize risk (strikes).

There are several questions to be explored with Spy TRACS, including: How does the player fuse his internal belief (pre-spy) with an external report (from spy) to establish his final belief? How does his trust in spies depend on the format and content of information display, and on the spy’s stated reliability (%) versus actual reliability (#correct reports/#total reports)?

Investigations

Thus far, our research has focused on the simplest (solitaire) game of Straight TRACS. Below we report on specific mental limits that we have identified in experiments with this game. These limits, which represent opportunities for decision support in practical applications, include: (1) Baseline Bias: where human subjects exhibit an initial bias in their mental representation of the baseline odds. (2) Anchoring and Adjustment: where human subjects exhibit a subsequent bias towards the baseline anchor as they make dynamic adjustments to the odds. (3) Concurrent Counting: where human subjects are limited to about 3 items per set when counting items in multiple sets. (4) Illusory Importance: where human subjects overestimate the benefit of a decision support system and misjudge the benefits of various coping strategies.

Testing TRACS

To “win” at TRACS, by choosing the “best” track on each turn, a player must update the baseline odds as the deck is depleted. Our first experiment, reported here (also see Burns 2002), was designed to measure how well people could perform this task. Our probe was a “confidence meter” that players set before each turn to indicate subjective belief in the tread (color) of each track. The confidence meter (Figure 1) displayed a discrete set of buttons on a scale ranging from 100% Red (far left) to 100% Blue (far right). Before each turn, the player clicked one of the buttons to report his subjective belief. There are 11 turns in each game, and each turn is a choice between two tracks, so each game yields 22 data points.

We tested 43 subjects on a personal computer using a mouse interface. Each subject played 10 games, and each of the 43*10=430 games used a random shuffle of the standard deck (Table 1). There were no time limits, but each game was typically completed in less than 5 minutes.

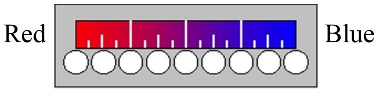

Figure 1: Confidence meter for measuring subjective belief.

The player's problem is illustrated in Figure 2. This figure shows the actual odds for a typical game, where odds are measured in % Red (% Blue = 100% - % Red). Figure 2 shows that the actual odds start out at their baseline values (see Table 1) of 75% Red for triangle tracks, 50% Red for circle tracks and 25% Red for square tracks. The actual odds then change (moving up/down on Figure 2) as tracks are turned and treads are revealed during the game (moving right on Figure 2).

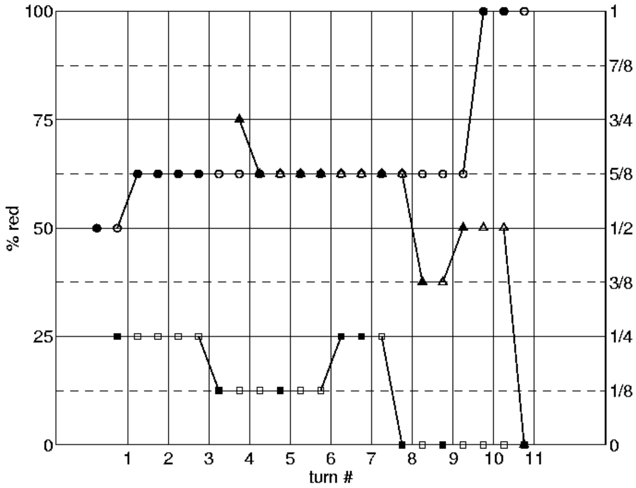

Figure 2: Actual odds (computed) for a typical game.

Figure 3 illustrates a typical player's performance in the same game. This figure shows the player's mental odds during the game, as reported by his setting of the confidence meter. Compared to the actual odds (Figure 2), we see that the mental odds (Figure 3) exhibit a bias towards the baseline (initial) odds. For example, after a minor adjustment near the start of the game, the player reported constant odds for circles even after the actual odds for circles had moved far from the baseline. The player did somewhat better for triangles, although still exhibiting a bias towards the initial baseline. The player did very well for squares.

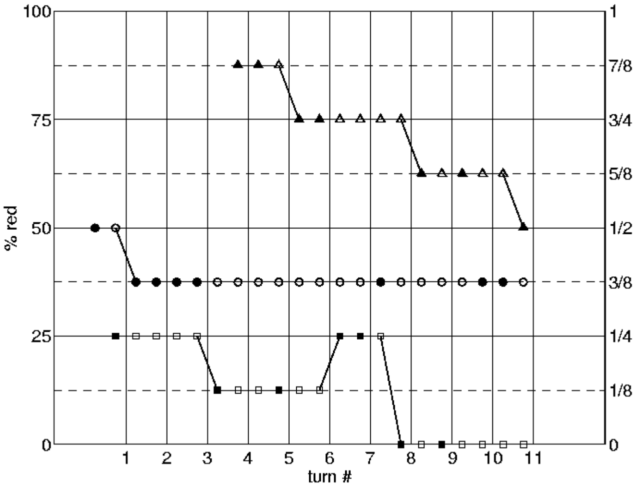

Figure 3: Mental odds (reported) for a typical game.

Baseline Bias

To choose the best track on each turn, a player must internalize the dynamically changing odds. In particular, he must be able to detect when there is an "inversion" (crossover) in the odds for two track types relative to their initial (baseline) configuration. For example, late in the example game (Figure 2), the player should choose a circle track rather than a triangle track to match a Red tread, since at this point in the game a circle is more likely than a triangle to be Red. The player (Figure 3) did not, i.e., he did not detect the odds inversion.

As a composite measure of human performance (Figure 4), we computed average error (all players, all games) versus turn in game. Error is defined as the absolute value of player odds (reported by the subject on the confidence meter) minus actual odds (computed from the cards remaining in the deck).

As seen in Figure 4, the error results for human subjects are similar to those of a "baseline agent". The baseline agent is a simulated player who played all 430 games using the baseline odds, i.e., by setting the confidence meter at the baseline odds, and choosing tracks accordingly, on every turn of every game. Figure 4 shows that human subjects performed, on average, slightly worse than the baseline agent at the start of games and slightly better than the baseline agent at the end of games.

It is interesting that the curve for human subjects starts slightly above the curve for the baseline agent. Since the actual odds are equal to the baseline odds at the start of the game, this means that people make errors in estimating the baseline odds themselves. We were rather surprised with this result because the baseline odds are graphically illustrated by the tread designs (on the fronts of the cards) and because there is always a tread on the field for the player to see.

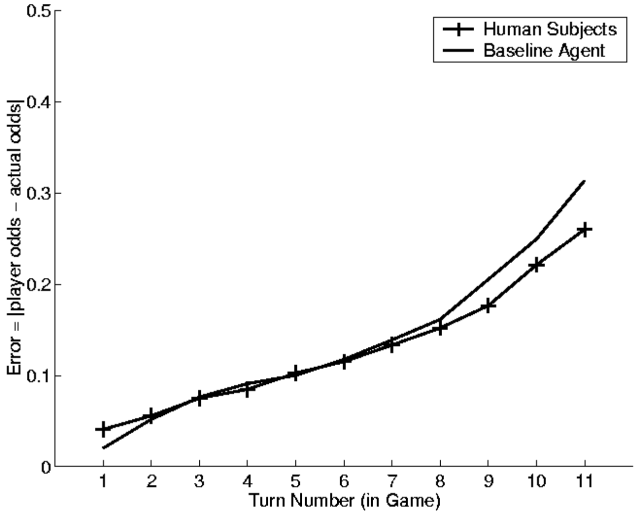

Figure 4: Average error versus turn in game.

To further explore this result, Figure 5 plots histograms of players' confidence settings (buttons along x-axis) for all "virgin" tracks. These are tracks encountered early in the game when actual odds are exactly equal to the baseline odds. The plots show that it is relatively common for players to be "one button too certain" for triangle and square tracks (but not for circle tracks). This finding is consistent with previous findings of an "overconfidence effect" in probabilistic reasoning (Gigerenzer et al. 1991). In our case, overconfidence appears as a bias in the mental representation of baseline odds that are >50% or <50%.

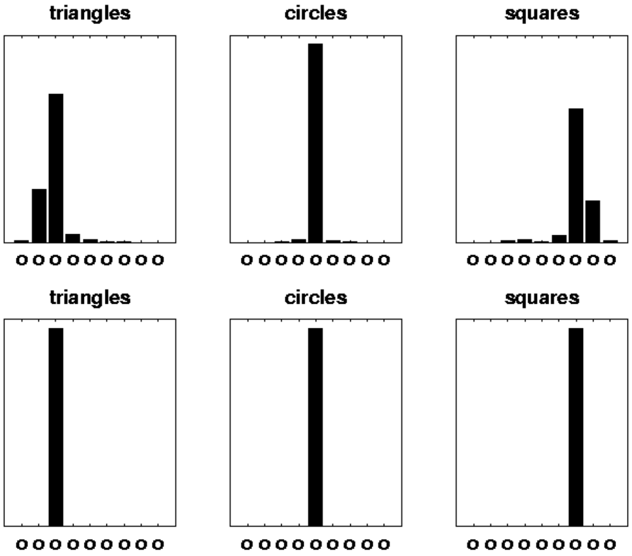

Figure 5: Bias in baseline odds. Top panels show subjects' (biased) response. Bottom panels show correct response. The x-axis represents buttons on the confidence meter.

Anchoring and Adjustment

The other bias that we see in Figure 4 is a subsequent bias, as the game proceeds, towards the baseline odds. That is, the error curve for human subjects remains close to that of the baseline agent compared to the optimal case of zero error (flat line at y=0). However, near the ends of games, Figure 4 shows that the error for human subjects is somewhat less than the error for the baseline agent. This means that people are not just playing the baseline odds, i.e., that they do have some success in updating odds. Overall, these results suggest that players are estimating the dynamic odds via a heuristic process of "anchoring and adjustment" (Tversky and Kahneman 1974).

To further explore the anchoring and adjustment, we examined how players' confidence meter settings change from one turn to the next for the same track type. We examined both the relative magnitude (Figure 6) and the qualitative type (Figure 7) of change. The qualitative types of change plotted in Figure 7 are defined in Table 2 relative to the baseline odds, which are the postulated anchor.

Figure 6 shows that human subjects made many more no-button adjustments (actually non-adjustments) to the confidence meter from one turn to the next, compared to an optimal agent who always sets the confidence meter at the actual odds. Figure 6 also shows that human subjects made many less one-button adjustments than the optimal agent. Figure 7 shows that human subjects made many more Type 0 adjustments and many less Type 1 and Type 2 adjustments than the optimal agent. These results clearly suggest that human subjects are making conservative "adjustments" to a baseline "anchor".

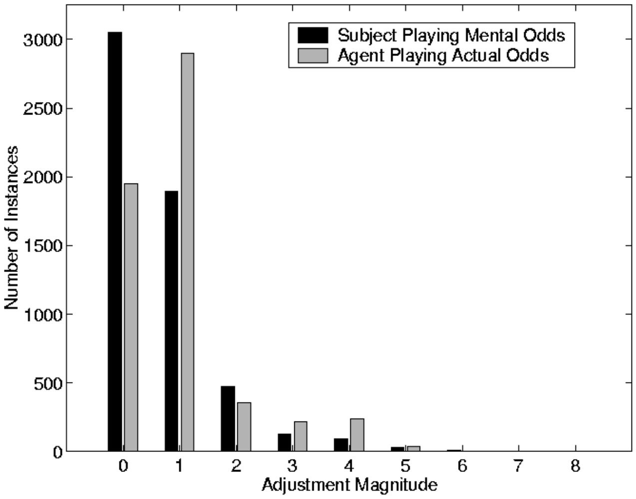

Figure 6: Number of adjustments by magnitude. Magnitude is measured in number of buttons on the confidence meter.

Table 2: Types of adjustments (anchor = baseline odds).

Type 0	Stay on anchor
Type 1	From on-anchor to off-anchor
Type 2	From off-anchor to more off-anchor
Type 3	From more off-anchor to less off-anchor
Type 4	From off-anchor to on-anchor

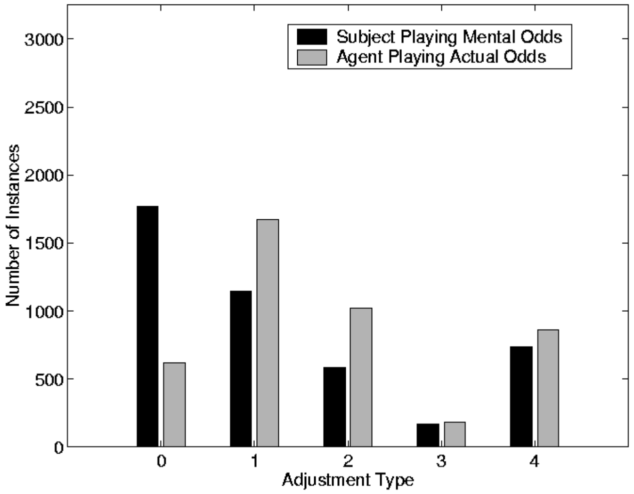

Figure 7: Number of adjustments by type. Types are defined in Table 2.

Concurrent Counting

The question, of course, is what (exactly) are the mental limits that prevent more accurate adjustments? Mathematically, accurate adjustments require concurrent counting of cards in each of 6 sets (see Table 1), then normalizing numbers to convert the counts to odds. For example, given baseline odds of 6/8 Red for triangle tracks, assume that 2 Red triangles and 1 Blue triangle have been revealed on earlier turns. Then, there are 6-2=4 Red triangles and 2-1=1 Blue triangle remaining in the deck, so the updated odds for triangles are 4/(4+1) = 4/5 Red.

To gain more insight into the mental process of concurrent counting, we examined how well players could count cards to odds of certainty. For analysis purposes, the 6 different card types (Table 1) are grouped into three different classes (2-count, 4-count, 6-count) that reflect how many cards of a single type must be counted to achieve certainty. For example, after counting 2 Red squares the odds for squares are 0% Red, and after counting 2 Blue triangles the odds for triangles are 100% Red. Thus, the 2-count class includes Red squares and Blue triangles. Similarly, the 4-count class includes Red circles and Blue circles, and the 6-count class includes Red triangles and Blue squares (refer to Table 1).

Figure 8 shows how well human subjects performed in making the 2-count, 4-count and 6-count during the experiment. This figure plots total "hits" (all players) of each n-count, where a hit means that a player indicated certainty correctly by clicking the leftmost or rightmost button on the confidence meter. A "miss" means that a player reported odds at one or more button(s) away from certainty when in fact the actual odds were certain. The results show that people are quite limited in concurrent counting of all three classes, and that there is a very large drop in hit rate between the 2-count and the 4-count.

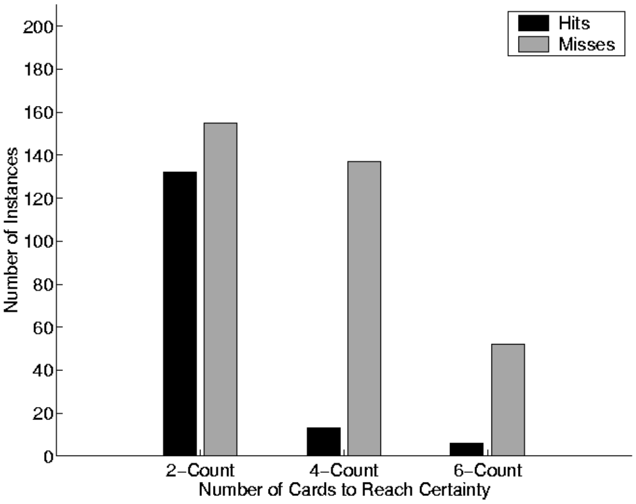

Figure 8: Detecting certainty by counting cards.

These results suggest that a basic limit of human performance in the task of concurrent counting is about 3 items/set. It is interesting to note that that a similar number (3±1) has been well established as a limit of "subitizing" in human visual perception (Dehaene 1992, also see Cowan 2001). In our case, the limit of 3±1 appears to constrain how many items/set an unaided mind can accurately enumerate in parallel.

Illusory Importance

With self-knowledge of one's limitations, an effective decision maker must allocate internal (mental) and external (system) resources to achieve desired outcomes in "bounded rationality". As such, we are interested in how well people can assess the benefits of various "coping strategies" that they might use and various "support systems" that they might buy.

Regarding a support system, we polled players to see how much they thought a perfect card-counting system would help in the game of Straight TRACS. The figure of merit was the average number of strikes per game that could be avoided (compared to the player's performance during the experiment) if the player had a computer system that displayed the actual odds at all times.

As seen in Figure 9, players think the support system would offer a considerable improvement. Most players (case c in Figure 9) believe it would save at least two strikes per game, and many players believe it would save more than two strikes per game. Since the average number of strikes per game is about 4 (for a baseline player with no card counting), a savings of 2 or more would be very significant.

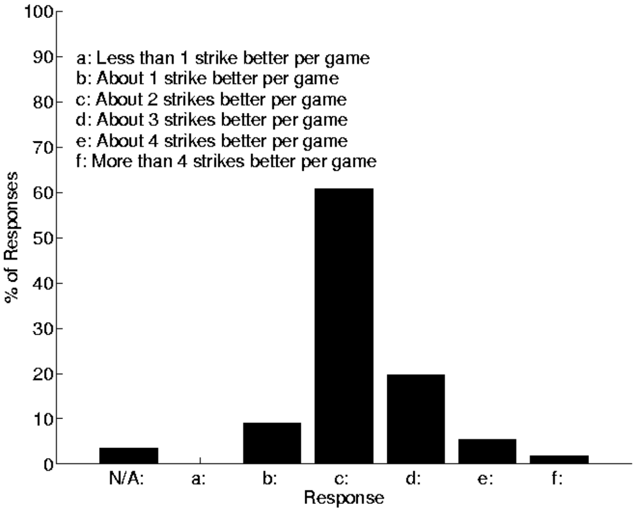

Figure 9: Benefit of perfect card counting (players' belief).

Contrary to subjects' belief, agent-based simulations show that the proposed system would save less than one strike per game (case a in Figure 9). This is a concrete example of how subjective beliefs can be biased by effects like "ease of recall", "illusion of control" and "wishful thinking" (Sage 1981). It is also an example of how simulations can improve subjective intuitions by providing objective demonstrations of a proposed system's benefit.

Besides support systems, training can also help people make better decisions. But, like the case of system designers, this requires that people (trainers) be able to judge the effectiveness of various "coping strategies" so that they can advise others (trainees) accordingly. Motivated by the popular notion that short term human memory is limited to about 7 items (Miller 1956), we asked players which set of 8 or less cards they thought it was most important to count in the game of Straight TRACS. That is, if players could not have a support system (to extend cognitive limits) but instead had to adopt a coping strategy (within cognitive limits), then what coping strategy would they adopt (and recommend to others) to minimize the risk of strikes?

As seen in Figure 10 (case d), most players think it is best to count the "rare" cards. Only about one quarter of all players said it would be best to count the circles (case c). To objectively assess these responses, we programmed software agents to use various strategies (Table 3) in the same games played by human subjects (who we polled).

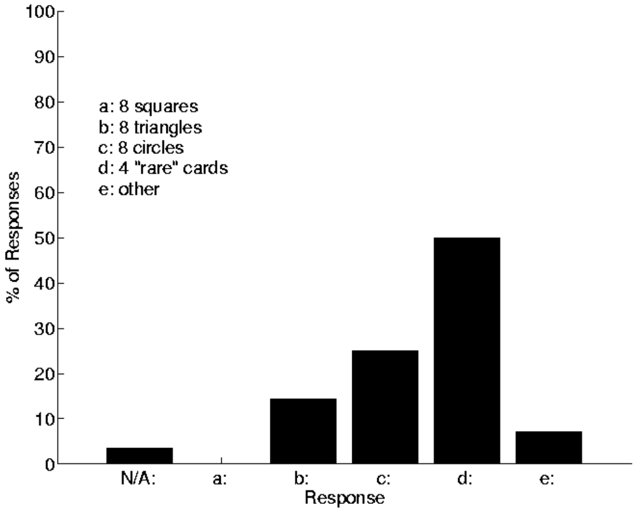

Figure 10: Most important cards to count (players' belief).

Contrary to the dominant belief (case d in Figure 10), Figure 11 (strategy 2) shows that it does little good to count the rare cards. Counting the circles (case c in Figure 10, strategy 6 in Figure 11) is much better. In fact, even counting circles of just one color (strategy 3) is much better than counting all of the rare cards (strategy 2).

Table 3: Various card-counting strategies (see Figure 11).

1	Count 0 cards (same as using baseline odds)
2	Count 4 rare cards (2 Red squares & 2 Blue triangles)
3	Count 4 circles (4 Red circles or 4 Blue circles)
4	Count 8 squares (2 Red squares & 6 Blue squares)
5	Count 8 triangles (6 Red triangles & 2 Blue triangles)
6	Count 8 circles (4 Red circles & 4 Blue circles)
7	Count 24 cards (same as using actual odds)

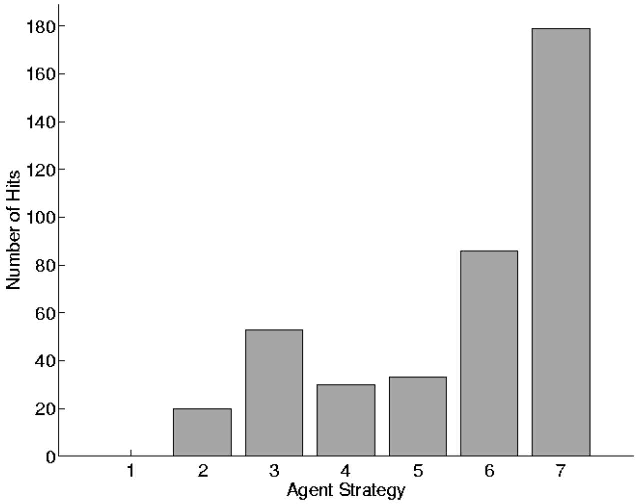

Figure 11: Benefit of limited card counting, computed from agent simulations. Agent strategies are defined in Table 3. "Hits" are instances where odds inversions are detected.

Figure 11 uses "odds inversions" to measure the benefits of various card-counting strategies. This is the appropriate figure of merit because, if there is no inversion, then baseline odds (with no card counting and no odds updating) can be used to choose the best track on each turn. Evidently people are using some other figure of merit when they judge the "importance" of various card types. This exemplifies the kinds of judgments for which simulations might help training, i.e., to help people understand when and why their subjective beliefs deviate from objective standards.

Applications

The mental limits discussed above, i.e., Baseline Bias, Anchoring and Adjustment, Concurrent Counting and Illusory Importance, represent opportunities for decision support in chance discovery and management applications. For example, an obvious support system for Straight TRACS is a card-counting system that always displays the current (actual) odds. Such a support system would be "classic" in the sense that it performs a function that is difficult for the unaided mind yet is easy to implement in a computational device.

One contribution of our research is to show that such support systems are not always what they are cracked up to be. That is, our simulations show that the benefit of the proposed system in the game of Straight TRACS is significantly less than players intuitively forecast. We suspect that the same is true for support systems proposed in many practical applications, and we suggest that computer simulation can be a valuable way to more objectively evaluate the benefits of such systems. Our research also illustrates the benefits of simulation for identifying coping strategies that are candidates for training, i.e., because they are effective yet not intuitive.

Of course, to evaluate the benefits of support systems (and coping strategies) via computer simulations, one needs software agents that can perform like the people that these systems are intended to support. This requires computational models of cognitive limits, of the kind that we are developing in our experiments. Thus, our research involves both experiments with people and simulations with agents in order to advance the practice of decision support engineering.

Finally, we note that our initial investigations (reported here) were specifically designed to address cognitive tasks in which people are known to have significant limits (e.g., counting cards and updating odds). However, people also exhibit significant powers of intuition and creativity in naturalistic decision making. Other versions of the TRACS game are specifically designed to tap these talents, in order to gain insight into the adaptive advantages of the human mind. A goal of our future investigations with these games is to develop computational models of cognitive strengths, so that artificial agents can be designed with useful features of natural intelligence.

Conclusions

We presented a new card game called TRACS that provides a micro world for research on chance discovery and management. Our initial investigations with the game have identified a number of mental limits that suggest practical opportunities for support systems and training. Future experiments and simulations will use TRACS to further explore human performance in tasks of Risk Assessment, Resource Management and Rational Engagement (in collaboration and competition). These tasks are the prototypical challenges of command and control in naturalistic decision making. As such, we believe that our micro world investigations with TRACS can inform real world applications in the field of decision support engineering.

Acknowledgments

This research was supported by the MITRE Technology Program. Thanks to Craig Bonaceto for help in performing experiments and programming simulations.

References

Burns, K. 2001a. TRACS: A Tool for Research on Adaptive Cognitive Strategies: The Game of Confidence and Consequence. On the web at www.tracsgame.com.

Burns, K. 2001b. Introducing TRACS: A unique game and a useful tool. Annual Meeting of the Society for Judgment and Decision Making. Orlando, Florida.

Burns, K. 2002. On Straight TRACS: A baseline bias from mental models. In Proceedings of the 24th Annual Meeting of the Cognitive Science Society. Fairfax, Virginia.

Cowan, N. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24: 87-185.

Dehaene, S. 1992. Varieties of numerical abilities. *Cognition* 44:1-42.

Gigerenzer, G.; Hoffrage, U.; and Kleinbölting, H. 1991. Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review* 98:506-528.

Klein, G. 1998. *Sources of Power: How People Make Decisions.* Cambridge, MA: MIT Press.

Miller, G. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63:81-97.

Sage, A. 1981. Behavioral and organizational considerations in the design of information systems and processes for planning and decision support. *IEEE Transactions on Systems, Man and Cybernetics* SMC-11, 9:640-678.

Tversky, A.; and Kahneman, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185:1124-1131.