Title
On Straight TRACS: A baseline bias from mental models

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 24(24)

ISSN
1069-7977

Author
Burns, Kevin

Publication Date
2002

Peer reviewed

On Straight TRACS: A baseline bias from mental models

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

Abstract

TRACS (Tool for Research on Adaptive Cognitive Strategies) is a family of games played with a special deck of two-sided cards (see www.tracsgame.com). TRACS has the advantage of being both mathematically tractable to theoretical analysis and psychologically relevant to practical applications. The simplest game, called Straight TRACS, is a series of choices where the player must turn over one of two cards to match a third card. The object is to make the most matches on a trip through the deck. The challenge is to track the changing odds in order to make the best choices. We performed experiments and simulations to measure human performance in this probabilistic and dynamic task. We present our finding of a Baseline Bias, in which subjective odds are (incorrectly) anchored to the baseline odds. This is an interesting result because it is contrary to other well-known biases, such as Gambler's Fallacy, in which subjective odds are (incorrectly) not anchored to the baseline odds. We propose a theory of mental models to reconcile our finding with previous research on heuristics and biases.

Introduction

A dilemma of decision research is to obtain the rigors of controlled experimentation yet maintain some relevance to practical applications. Our approach is a synthetic task environment (Gray, in press) called "TRACS" (Burns, 2001a) that replicates the cognitive challenges of naturalistic decision-making (Klein, 1998), including probabilistic risk assessment and dynamic resource allocation.

TRACS is both a unique game and a useful tool (Burns, 2001b). From a mathematical perspective, TRACS is unique because it is played with a special deck of two-sided cards, and because it has extensible rules that allow the same game to be played alone or with others.

Unlike standard playing cards, the backs of the cards provide clues to the fronts, and the deck contains unequal numbers of each card type (Table 1). Compared to Poker and other games of imperfect information, the two-sided cards make TRACS more tractable to theoretical analysis of optimal solutions. Compared to Chess and other games of perfect information, the two-sided cards make TRACS more relevant to diagnoses and decisions in practical domains like business, medicine and warfare.

From a psychological perspective, TRACS is useful because it provides a naturalistic micro world for experiments and simulations. Unlike other approaches to research on probabilistic reasoning, which often employ verbal stimuli in the form of static questions, TRACS employs graphical stimuli in a game of dynamic situations. This reduces artificial framing effects (see Nickerson, 1996) and introduces realistic temporal context.

We are using TRACS to perform experiments on human subjects and to perform simulations with software agents. Our experiments are designed to elicit cognitive strategies and our simulations are designed to evaluate these strategies against normative standards. Taken together, our experiments and simulations allow us to build and test models of cognitive competence that are relevant to practical applications in command and control (Burns, 2001c).

This paper reports on our first experiment and simulations using the simplest version of the game, called Straight TRACS. We explain the game, discuss our experiment and present our finding of a Baseline Bias. We also propose a theory of mental models to reconcile our finding with previous research on heuristics and biases.

The Game

Straight TRACS is a solitaire game played with 24 two-sided cards (Table 1). The backs of the cards, called "tracks", show black shapes (triangle, circle or square). The fronts of the cards, called "treads", show colored sets (Red or Blue) of these same shapes. Table 1 shows the distribution of shape/color (track/tread) cards in the deck. This distribution defines the baseline odds. For example, at the start of the game, a triangle track is likely (6/8 = 75%) to be Red, a square track is likely (6/8 = 75%) to be Blue and a circle track is 50-50. However, during the game, the odds change as the deck is depleted.

Table 1: Distribution of cards in the deck. The backs are called "tracks" and the fronts are called "treads".

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

To play Straight TRACS, 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's task is to turn over one of the two tracks (revealing 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 clash. The two treads are removed from the field and the remaining track is turned to reveal its tread. This becomes the tread to match on the next turn. Two new tracks are dealt from the deck, a track is turned, the treads are scored, etc. Play continues until all cards (except the last two, which do not count) have been paired and scored. The object of the game is to minimize strikes on a trip through the deck.

Experiment

The goal of our experiment was to measure how people track the changing odds in TRACS. The probe in our experiment was a confidence meter that players set before each turn to indicate their subjective belief in the color (tread) of each shape (track) on the field. We used two different confidence meters (Figure 1), both based on a spectrum that runs from 100% Red to 100% Blue. One confidence meter displayed a discrete set of qualitative values on an octal scale. The other confidence meter displayed a continuous set of quantitative values on a decimal scale.

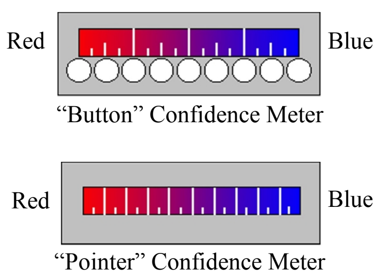


Figure 1: Two different confidence meters.

We tested 43 adults playing 10 games each. Subjects were tested on a personal computer using a mouse to set the confidence meter. There were no time limits, but each game was typically completed in less than 5 minutes. Each subject played in two blocks of 5 games, one block with each confidence meter in balanced design to control for fatigue and learning effects. The two blocks were separated by a short break. Before data collection, subjects read a playbook that described the cards and rules, watched a demo and played a practice game. All games were played with random shuffles of the deck and all 43*10 = 430 shuffles were unique. The experimental results were similar for Button and Pointer confidence meters, so all data is combined here, rounding Pointer data to the nearest Button for consistency.

Analysis

Baseline Bias

The player's problem is illustrated in Figure 2, which shows the actual odds for a typical game. By convention, we measure odds in % Red, where % Blue = 100 - % Red. Figure 2 shows that the odds for each track type start at their baseline values (75% Red for triangles, 50% Red for circles and 25% Red for squares). However, the actual odds change (moving up/down on Figure 2) as tracks are turned to reveal their treads (moving right on Figure 2).

Figure 2: Change in actual odds (typical game).

Figure 3 illustrates a typical player's solution to the problem illustrated in Figure 2, as reported by the player's setting of the confidence meter for each track (before each turn). Relative to the actual odds (Figure 2), we see that the reported odds exhibit a bias towards the baseline odds. For example, after a minor adjustment near the start of the game, the player (Figure 3) reported constant odds for circles even after the actual odds (Figure 2) had moved far from the baseline. This Baseline Bias is explored further below.

Odds Inversions

Recall that the object of Straight TRACS is to turn the track that is most likely to match a given tread. At the start of the game, this is a simple task since the baseline odds specify which track to turn, e.g., triangle rather than circle to match Red. However, as the deck is depleted, the actual odds for two track types may become "inverted" with respect to their baseline configuration. This occurs whenever there is a crossover of two track types on the dynamic odds plot (Figure 2).

Figure 3: Change in reported odds (typical game).

For example, in Figure 2, a crossover near the middle of the game causes the odds to be less % Red for triangles than for circles for the remainder of the game. This is an inversion of the baseline odds relation between triangles and circles. Figure 3 shows that the player failed to detect this odds inversion.

As a gross measure of cognitive competence, we treat each odds inversion as a signal that a player must detect in order to minimize strikes in Straight TRACS. Figure 4 shows the total number of hits, misses and false alarms for this signal (for all players and all games). The relatively small number of hits compared to misses demonstrates that subjects exhibit a Baseline Bias. The occurrence of some hits and false alarms suggests that, although biased towards the baseline odds, subjects are at least trying to update odds, i.e., they are not just playing the baseline odds.

Figure 4: Detection of odds inversions.

Average Error

Each odds inversion (see above) involves a pair of tracks. As another measure of cognitive competence, we also examine confidence errors for single tracks. Figure 5 shows the average error versus turn in game, for human subjects and for a simulated agent that always plays the baseline odds.

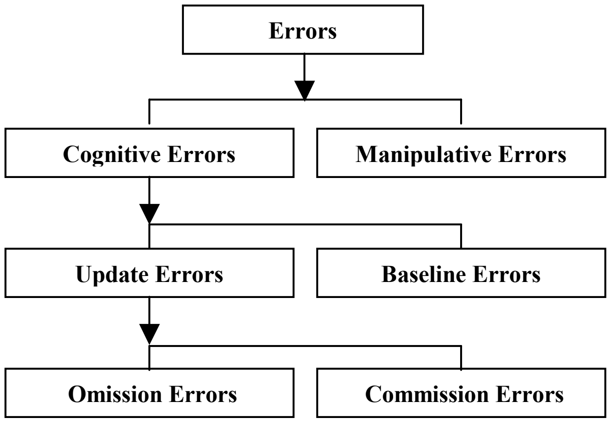
Figure 5 shows that error increases with turn, i.e., as the actual odds deviate more from the baseline odds, for both the human subjects and the baseline agent. This shows that people have a Baseline Bias relative to the actual odds (zero error). Figure 5 also shows that the average error is higher for human subjects than for the baseline agent at the start of the game. This is a surprising result because: (1) The baseline odds are explicitly illustrated on the cards (treads) for the player to see. (2) The actual odds are obviously equal to the baseline odds at the start of the game. (3) Playing the baseline odds is a strategy that requires virtually no mental effort.

Figure 5: Average error in reported odds.

Kinds of Errors

To help explain Figure 5, we define a taxonomy of errors (Table 2). We first distinguish between Cognitive Errors, which are mental errors in judging odds, and Manipulative Errors, which are physical errors in moving the mouse to match the mind. We then distinguish between two kinds of Cognitive Errors: Update Errors are mental errors in updating odds relative to baseline odds, and Baseline Errors are mental errors in estimating the baseline odds themselves. Finally, we further distinguish between two kinds of Update Errors: Omission Errors are where no mental update is performed when it should be, and Commission Errors are where a mental update is performed but the result is incorrect.

Table 2: A taxonomy of errors.



The baseline agent makes no Manipulative Errors, no Baseline Errors and no Commission Errors, i.e., it makes only Update Errors of Omission. In fact, since the baseline agent never updates odds, its performance provides an upper bound on the magnitude of Omission Errors. Figure 5 shows that the baseline agent's Update Error is non-zero on turn 1. This is because the tread on the first field can cause a change in odds before the first turn. For example, assume that the cards on the first field are circle (track), Red circle (tread) and square (track). The baseline odds for circles are 4/8 Red, but since one Red circle is revealed as a tread on the field, the actual (updated) odds for circles are 3/7 Red. The same effect is magnified on later turns as more treads are revealed, hence the Omission Error increases with turn (Figure 5, curve for baseline agent).

For human subjects, the total error comprises Manipulative Error, Baseline Error and Update Error (Omission and Commission). The difference between total human error and baseline agent error on turn 1 is attributed to Manipulative Error and Baseline Error, which we assume are relatively independent of turn in game. Thus, the curve for total human error can be shifted downwards (curve * in Figure 5) to get an estimate of human Update Error.

This shifted curve for human error is directly comparable to the error curve for the baseline agent, which also includes only Update Error. The comparison (Figure 5) shows that human subjects are biased towards the baseline agent, relative to the actual odds (zero error). However, the shifted curve also shows that human subjects outperform the baseline agent (who does not update odds), and the difference grows with turn as the difference between actual odds and baseline odds increases. Thus, we conclude that human subjects make fewer Omission Errors than the baseline agent, and that the Commission Errors made by human subjects are not significantly larger in number and magnitude than the baseline agent's Omission Errors.

Anchoring and Adjustment

These results are consistent with the well-known heuristic strategy of "anchoring and adjustment" (Tversky & Kahneman, 1974). In our case, the baseline odds (with some Baseline Error, see above) are the anchor to which people make adjustments. The adjustments are, on average, better than pure anchoring but significantly worse than optimal adjusting.

To gain further insight into the adjustments, we examine how the confidence meter settings change from turn to turn (for the same track type). We define various magnitudes of adjustment (i.e., no-button jump, one-button jump, two-button jump, etc.) and compute the number of times each magnitude of adjustment was made. Figure 6 compares the results for human subjects to an optimal agent (playing the same games) who always sets the confidence meter at the actual odds. As expected from anchoring, human subjects most often make a no-button adjustment. This is in contrast to the optimal agent, who most often makes a one-button adjustment.

Figure 6: Number of adjustments (by magnitude).

Besides the magnitude of adjustment, we also examine various types of adjustments. We define five types of adjustments (Table 3) and compute the number of times each type of adjustment was made. Figure 7 compares the results for human subjects to an optimal agent (playing the same games) who always sets the confidence meter at the actual odds.

Table 3: Types of adjustments (anchor = baseline odds).

Type 0	No adjustment
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

As expected from anchoring, Figure 7 shows that human subjects make many more Type 0 adjustments (actually non-adjustments) and less Type 1 and Type 2 adjustments than the optimal agent. Figure 7 also shows that the difference between Type 1 and Type 4 adjustments is smaller for human subjects than for the optimal agent. This suggests that, when people do move off the baseline anchor (Type 1) in an attempt to adjust odds, they often "lose it" and return to the baseline anchor (Type 4). The optimal agent moves off the baseline anchor (Type 1) more often and returns to the anchor (Type 4) only when the actual odds are equal to the baseline odds (i.e., the agent never "loses it").

These results (Figures 6 and 7) further support our conclusion that Baseline Bias in TRACS (Figures 4 and 5) is caused by a heuristic strategy of anchoring and adjustment.

Figure 7: Number of adjustments (by type).

The question, of course, is how (exactly) do people decide when and how much to move off anchor? That is, what (exactly) are the mental limits that prevent more accurate adjustments? The answer is crucial if we are to explain and predict human performance in TRACS or any other domain where people must think probabilistically about things that are changing dynamically. Below we propose a theory of mental models that takes a first step in this direction.

Theory

Mind Sets

We claim (Burns, in press; 2002; 2001b; 2001c; 2001d) that people make sense of the world by forming probabilistic models in their heads (see Knill & Richards, 1996; Johnson-Laird, 1994; Gigerenzer et al., 1991).

We further claim that: (1) Mental models are bounded by natural regularities of the world. (2) Mental models have a normative basis within their natural bounds. (3) Cognitive competence can be explained and predicted by specifying the natural bounds and normative basis of mental models (i.e., in bounded rationality, see Simon, 1990).

More specifically, we claim that probabilistic mental models can be characterized as computational "mind sets" (Burns, 2002). Like the term "model", which is both noun and verb, we use the term "set" in dual sense to refer to both the mental representation of classes (e.g., declarative knowledge) and the mental operation of routines (e.g., procedural knowledge). The central tenet of our theory (Burns 2002) is that these mind sets are normative within their cognitive bounds.

As an initial test, below we sketch how our theory can explain Baseline Bias in TRACS. We also sketch how our theory can reconcile Baseline Bias with previous findings of Gambler's Fallacy and Base Rate Neglect in other probabilistic reasoning tasks (Tversky & Kahneman, 1974). This is a non-trivial test of the theory, since Gambler's Fallacy and Base Rate Neglect appear at first glance to be contrary to Baseline Bias.

Gambler's Fallacy and Base Rate Neglect

In Baseline Bias (in TRACS), people do not update the baseline odds when they should. Conversely, in Gambler's Fallacy, people update the baseline odds when they should not. Furthermore, in Base Rate Neglect, people discount or ignore the baseline odds altogether. How can we explain these differences? According to our theory, all three biases occur because people reason about probabilities with mind sets.

For Base Rate Neglect (Tversky & Kahneman, 1974; Koehler, 1996; Cosmides & Tooby, 1996), we suggest that people ignore the baseline odds in light of other evidence because they believe that the baseline odds reflect a less relevant (not applicable) set of occurrences. It is difficult in theory, let alone in practice, to aggregate probabilities that are derived from diverse sources with different pedigrees. As such, it is a bounded-Bayesian strategy to rely on the one source that is judged to be most relevant and reliable. Base rates that are judged irrelevant or unreliable are therefore neglected.

For Gambler's Fallacy (Tversky & Kahneman, 1974), we suggest that people update the baseline odds because they are judging odds for a finite (bounded) set rather than for an infinite set. For example, after seeing 10 heads and 2 tails, a gambler who believes the coin is fair will think that the future holds more tails than heads, simply because he thinks that the eventual (large but finite) set of many tosses for this coin will be balanced. As such, it is a bounded-Bayesian inference to conclude that the future odds are slightly higher for tails than for heads.

For Baseline Bias (in TRACS), we suggest that people want to update odds (as they tell us) but that it is simply beyond their cognitive capacity. To do so, players must count cards in each of six sets (see Table 1) and normalize to convert the counts to odds. These two tasks, i.e., concurrent counting and normalizing numbers, are naturally hard for the unaided mind (Dehaene, 1997; Dehaene, 1992; Gallistel & Gelman, 1992; Nickerson, 1996). Thus, with self-knowledge of mental limits, it is a bounded-Bayesian strategy to remain anchored to the baseline odds unless and until the evidence for an adjustment is compelling. For example, in the extreme case, pure anchoring to baseline odds (i.e., never adjusting) is the bounded-Bayesian strategy for a decision maker who knows that he cannot remember which cards have been revealed in the course of a game.

Conclusion

Our initial experiment and simulations show that TRACS provides a useful micro world for investigating how people make diagnoses and decisions under uncertainty. Our finding in Straight TRACS is that players exhibit a Baseline Bias, which we attribute to a heuristic strategy of anchoring and adjustment. We sketched a theory of set-based mental models that reconciles our finding with previous research on heuristics and biases. Our future plans are to use TRACS to investigate the mental limits of concurrent counting, normalizing numbers and other facets of cognitive competence in probabilistic and dynamic reasoning.

Acknowledgements

This research was supported by the MITRE Technology Program. Thanks to Craig Bonaceto, Eric Forbell and Fritz Behr for their work on the experiment and simulations.

References

Burns, K. (2001a). TRACS: A Tool for Research on Adaptive Cognitive Strategies: The Game of Confidence and Consequence. *Published on the World Wide Web,* 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. (2001c). A Bayesian basis for mental models. Draft manuscript.

Burns, K. (2001d). Mental models of line drawings. *Perception*, 30, 1249-1261.

Burns, K. (2002). Mind sets at TRACS: How people deal with changing odds. *40th Annual Bayesian Research Conference*, Studio City, California.

Burns, K. (in press). Mental models and normal errors. In Brehmer, B., Lipshitz, R., Montgomery, H. (Eds.), *How Do Professionals Make Decisions?* (Lawrence Erlbaum). Also presented at *5th Conference on Naturalistic Decision Making*, Tammsvik, Sweden, May 2000.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1-73.

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

Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics* (New York: Oxford).

Gallistel, C., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44,* 43-74.

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

Gray, W. (in press). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments and micro worlds in basic and applied cognitive research. *Special Joint Issue of Cognitive Science Quarterly and Kognitionswissenschaft*.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition, 50,* 189-209.

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

Knill, D., & Richards, W. (1996). *Perception as Bayesian Inference* (New York: Cambridge University Press).

Koehler, J. (1996). The base rate fallacy reconsidered: Descriptive, normative and methodological challenges. *Behavioral and Brain Sciences, 19,* 1-53.

Nickerson, R. (1996). Ambiguities and unstated assumptions in probabilistic reasoning. *Psychological Bulletin, 120,* 410-433.

Simon, H. (1990). Invariants of human behavior. *Annual Review of Psychology, 41,* 1-19.

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