AN ANALYSIS OF THE COMPUTABILITY OF HORSE RACES

by

TERESA SUSAN HOHOL

SUBMITTED TO THE DEPARTMENT OF
ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
IN PARTIAL FULFILLMENT OF THE REQUIRMENTS FOR
THE DEGREES OF

BACHELOR OF SCIENCE IN ELECTRICAL ENGINEERING

and

BACHELOR OF SCIENCE IN COMPUTER SCIENCE AND ENGINEERING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1985

Signature of Author_____
Department of Electrical Engineering and Computer Science
May 7, 1985

Certified by_____
Professor Joseph Ferreira Jr.
Thesis Supervisor

Accepted by_____
Professor David Adler
Chairman, Department Committee

An Analysis of the Computability of Horse Races

by

Teresa Susan Hohol

Submitted to the Department of Electrical Engineering
and Computer Science on May 10, 1985 in partial fulfillment
of the requirements for the Degrees of
Bachelor of Science in Electrical Engineering, and
Bachelor of Science in Computer Engineering

Abstract

The science of thoroughbred handicapping has often been an interesting and frustrating challenge. For any horse in a race, there are many pieces of data which may be used to predict its performance. The problem lies in the analysis of this large amount of seemingly unrelated data, and then in the comparison of the horses competing in the race. If a computer could be used to study the relevant data, and the factors involved, it would be possible to develop some models for the relationsnips which exist in this data. These models could then used in an analysis of the horses in a horserace to predict the winner of the race.

A computer system, called the WHINNY system, was developed to undertake this investigation. The WHINNY system consisted of two parts: an automated input system which allowed the large amount of necessary data to be entered mainly through the use of a pointing device, and an analysis system, which could be used to create models for relationships between the factors, and then "handicap" the race to predict the winner.

Several models were constructed using the WHINNY system. These models, when tested on actual horse races, were able to select the winners of two out of three horseraces. In addition, the results from these tests provided insight into some of the ways in which better models could be made in the future.

Thesis Supervisor: Dr. Joseph Ferreira, Jr.

Title: Associate Professor of Urban Studies
and Operations Research

Acknowledgments

The author wishes to thank the following people for their timely assistance and unwaivering support throughout this project:

Dr. Joseph Ferreira, Jr., the thesis supervisor, for his guidance, inspiration, motivation, and criticisms which helped to keep this project "winning".

Mr. Patrick A. O'Donnell, Research Engineer - M.I.T. Artificial Intelligence Laboratory, for patience and time he took to explain the care and feeding of the *Symbolics 3600* Lisp Machine, which was an integral part of this project.

Her parents, Mr. and Mrs. Marion H. Hohol, and sister Kathleen Hohol for their support, both moral and monetary, and their aid in collecting the data on which to test this project.

I would also like to thank, Miss Leova Wolf, secretary for Dr. Ferreira, and all of the people at the M.I.T. Artificial Intelligence Laboratory, for their friendship and advice throughout the duration of this project.

CONTENTS

FIGURES

TABLES

1. Introduction and Primer

Throughout the last few decades, computers have been used to solve many problems which were either too complex to handle previously, or just to tedious to merit effort. One such problem is the art of thoroughbred handicapping. The amount of information available on any horse entered in any race easily exceeds three hundred individual facts. When this is compounded by the fact that, on the average, a race consists of eight horses, and a race track holds ten meets per day, the amount of data available for analysis swells over twenty-four thousand pieces of data. This undoubtedly qualifies as tedious.

Thus the problem of thoroughbred handicapping appears to be a prime candidate to yield to the advances of computer power and technology. However, the massive amount of vital information required to conduct a reasonable handicapping analysis immediately presents two important problems: How to enter this much data, (much of which is repetitive in nature), and what to do with it once you have a handle on it. The first problem would be easy to solve if there were some way to obtain the data in a computer-readable format, or to read it off a paper using an optical scanning device. Unfortunately, the publishers of this data do not make it available other than in newpaper form, and the information is publishied in several different fonts (character sizes), which make it impossible to enter using an optical font reader. Thus, the only way to get this information into a computer is to enter it by hand. Answers can be found, however, for the second problem. There are a number of respectable volumes published on the art and science of thoroughbred handicapping, many of which will reap profits for those individuals patient enough to wade through the the tedium of crunching through the necessary analytic steps to determine the best horse (or the best bet). Unfortunately, many weekend horseplayers, myself included, have neither the time nor the patience to engage in such an analysis. There had to be a better way.

Certainly over the past few years, attempts have been made to find a better way. Various individuals have developed handicapping systems and betting systems, available at reasonable prices anywhere from five dollars, to five hundred dollars. *[1]* Some of systems are mere gimmicks; others make an honest attempt at trying to solve the tedium at producing valid bets. The approach which most, if not all of these systems take is to pick out a few (perhaps even key) factors from among the multitude, and to merely analyze those, at the cost of ignoring other vital data. The results of these systems are not encouraging. The few horseplayers I know who invested in one or more of these systems, lost both their investment, and subsequent dollars at the local track. There must be a better way.

One of the first serious efforts at applying computer power to the analysis of thoroughbred racing was carried out by Dr. William L. Quirin, Ph.D., and Associate Professor of Mathematics (Computer Science) at Adelphi University. Dr. Quirin studied the races of over five thousand horses during a season in New York, and statistically analyzed every possible factor known or thought to have some bearing on the results of a thoroughbred race. In his subsequent publication, *Winning at the Races* published in 1979, he divulged his findings, in the form of separate charts, one for each handicapping factor, and the percentage of winner found with factor, along with how much profit could have been made while betting only horses with this factor. The results supported many popular handicapping theories, and turned up some other unsuspected ones. Dr. Quirin also published in this book two computer-generated regression formulae, which he (and his computer) had developed, using those principles which his analysis had shown to be most effective. [Qurn79] His work certainly appeared to be a breakthrough for all horseplayers serious in their art, and tired of the tedium of handicapping.

While Dr. Quirin's analysis of past races proved to be a valid starting place for handicapping analysis, his regression formulae did not prove to be an end unto themselves. While the formulae may be accurate in their predictions, they proved to be unwieldy to handle, and very tedious to use. In addition, they did not take into account some of the tests which his analysis had "proved" to be effective, but which had been omitted, nevertheless, from the recursion formulae due to the excessive calculations necessary to carry them out. This however, would not be a problem, if the necessary calculations were carried out by a computer.

After studying the various approaches to thoroughbred handicapping and the results of Dr. Quirin's analysis, I became convinced that a systematic computer analysis of the available information about the entries of horse race, using those handicapping "angles" which had been proven effective by Quirin and others, would enable me to select the "best" horse out of enough races to make a profit. This would support the popular theory held by many successful handicappers that "You can't beat a race, but you can beat the races!"

In the following sections, I will first present a quick guide to thoroughbred racing, which will discuss briefly the terms involved, and how to read a summary of a horse's past performances, and what it all means. Next I will present a short review of the past efforts that various individuals have made to show that the races are beatable, and how my method of attack fits into the progress made in this area. Then comes my methodology; just what steps I took to organize the problem, and how I undertook to analyze horse races in detail. After this are the specific rating schemes I chose to compare the horses under analysis, along with how they work, and why I chose them. Following this is an explanation of how I build my analyses (of different combinations of the rating schemes), and how they go about comparing horses, using the rating schemes. Finally, I will present two short chapters, the first on how my computer analysis scored when applied to real races, and the second on what additional steps and

tests should be added to my analysis in order to increase it's performance.

Thoroughbred Racing Primer

This section is a brief description of thoroughbred racing, and how to understand past performances, and pari-mutual betting. *[2]*

For those of you who know the difference between "breezing" and "handily", or "win", "place", and "show", please go on to Chapter 2.

Now for the rest of you, let me unveil the glorious mysteries of the language and terminology behind the "Sport of Kings"! First, some of the more common terms of the business:

Thoroughbred - "purely bred", a breed of horse specially designed for speed and stamina. Somewhat more temperamental and high-strung than other horses, the thoroughbred is characterized by it's sleek, muscular look.

Thoroughbred race (or Flat race) - a race between thoroughbreds on flat tracks (no jumps), usually oval in shape, on either dirt or grass (turf), at distances from 5 furlongs to 2 miles. A furlong is an eighth of a mile. In a thoroughbred races, the jockey (rider) rides the horse, in contrast to harness races, where the driver rides in a sulky (2 wheeled cart) behind the horse. Also in a thoroughbred race, the horse is running at a gallop (the fastest gait), where in a harness race, the horses either trot or pace (slower gaits).

Thoroughbred Handicapping - the art and science of studying the horses entered into a thoroughbred race, and determining the horse best suited to winning. (or coming in second or third). People who conduct such analysis are deemed "Handicappers" (if they're successful enough!). Race tracks employ "track handicappers" to pick horses of similar abilities to enter into a race, so that no horse has a clear advantage over it's peers. (Our job is to find that horse!!)

Pari-mutual betting - the system where the money bet on a horse determines the horse's odds and subsequent payoff, should the horse win. More on this in a later section!!

In the Money - The condition of having come in first, second, or third in a race. *[3]*

Win - When one bets a horse to win, the horse must come in first for the horse to pay off.

Place - When a horse is bet to place, the bettor collects if the horse comes in either first *or* second.

Show - When a horse is bet to show, the bettor collects if the horse comes in either first, second, *or* third.

Angle - a characteristic or quality or test which a handicapper uses to base his decision that one horse is a better bet than another. For example, "Bet only horses which have won 2 races out of their last 5 starts" is one angle which many handicappers use. "Bet only black spotted horses when the moon is full" would be another angle, but probably not a very good one!

The Daily Racing Form - a daily newspaper which lists the horses running on that day at certain race tracks. Different papers are distributed for different parts of the country. The Eastern version covers tracks from Maine to Florida. The *Form* provides for each horse a recent history of the horse's last eight races, and any recent workouts that the horse has publicly done. horses running in each race that day. It also lists the jockeys who will be riding the horses today, and the weights that they will be carrying. In addition, the *Form* employs handicappers who try to pick the winners (and the second and third horse) of each race.

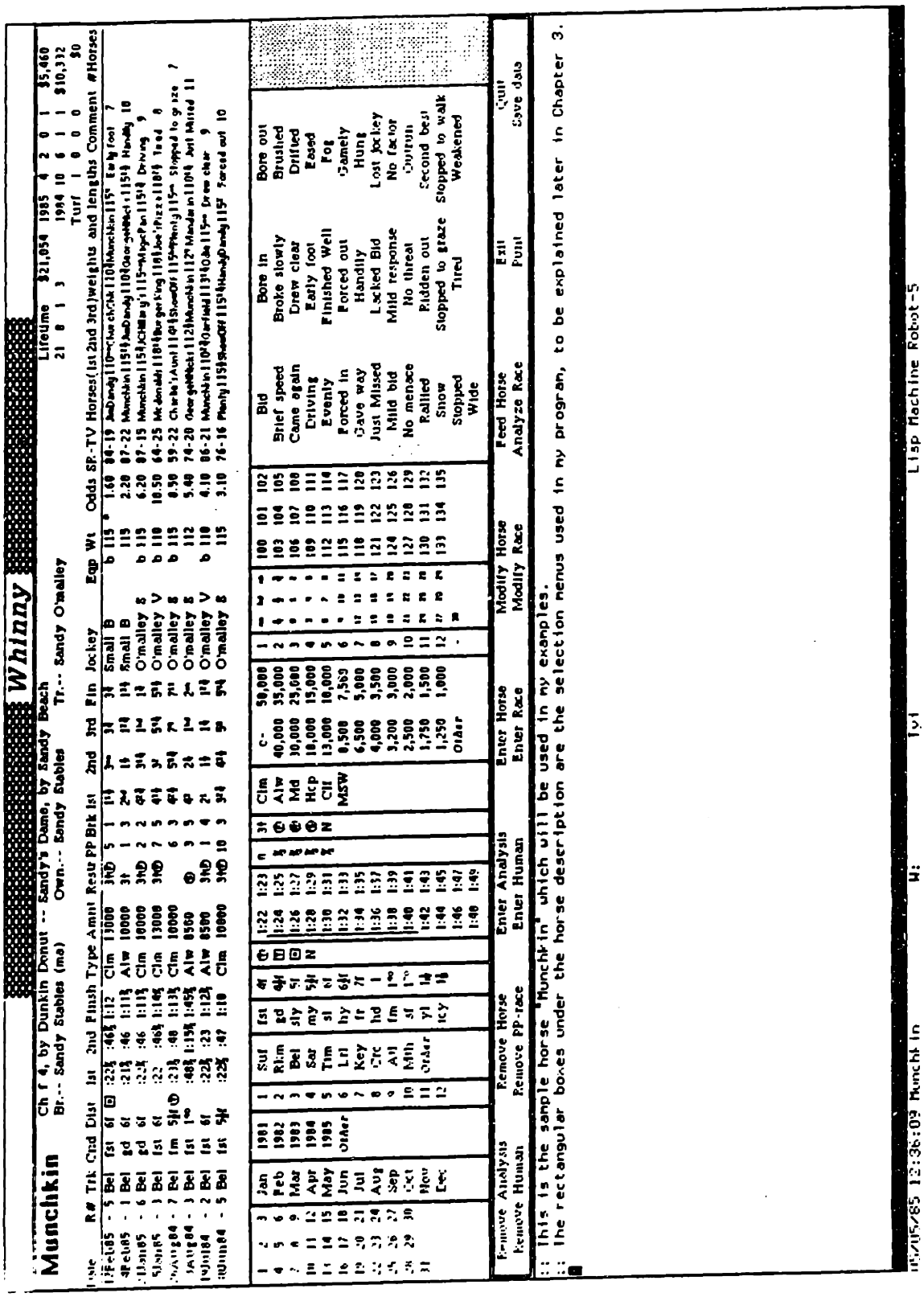
Past Performances - the detailed accounting of how the horse ran the last eight races it was entered in. Information in each past performances line include the date of the race, the track, the race number, the track conditions, fractional times for the race, the position of the horse under consideration at various points of the race, the jockey, the weight carried, and the horses which came in first, second, and third. Handicappers base much of their final decision on analysis of the horse's past performances.

How to read a horse entry and past performances

The *Daily Racing Form* publishes information about horses in a very precise, exact, terse manner, using symbols and abbreviations that can baffle and bewilder the novice handicapper. Once the meaning and interpretation behind the symbols is understood, however, reading horse information displayed in this format is both easy and meaningful. I have chosen to use this format of displaying horse information in my thesis, also, due to it's precise nature. (The fact that this is the form in which the data already comes in the *Form* doesn't hurt either!)

In the following pages, I will describe how to read this encoded information and also comment on the value it might have to a handicapper. After each description of a piece of data, I will give the corresponding piece of information from the sample horse "Munchkin", which can be found on the following page, Figure 1.

Figure 1: Sample horse Munchkin



The first part of a horse entry deals with information which is relevant mainly to the horse as an animal. They are the horse's :name, color, sex, age sire, dam, and grandsire.

Name - The first item of interest is the horse's name. It is located in the upper left hand corner of the horse entry, in large bold-face letters.

For our example: Name - Munchkin

Color - The next item is the color of the horse. This is useful in ir : the horse from it's peer at the racetrack (or at least narrow the possibilities). The colors for a thoroughbred are abbreviated as follows: bay (b.), brown (br.), black (blk.), chestnut (ch.), gray (gr.), roan (ro.), and white (w.)

For our example: Color - ch (chestnut)

Sex - Following the color, is the age of the horse. These are abbreviated as: colt (c.) (a male thoroughbred under age of five), horse (h.) (a male, five or older), filly (f.) (female, under five), mare (m.) (female, five or older), and gelding (g.) (a castrated male).

For our example: Sex - f. (filly)

Age - Following the sex, is the age of the horse. This is given in numbers, almost always between 2 and 9. Thoroughbreds all have the same official birthday, January 1st, so a horse born in December would be "1" (a yearling) in January. (and at an extreme disadvantage against his "same age" peers, the rest of his life). The age of a horse is often considered to be a deciding factor, as horses three or under are physically immature, and at a disadvantage when racing against older animals.

For our example: Age - 4

Sire - The next piece of information is the sire, the male parent of the horse in question. As many abilities important to racing ability, such as stamina, or the ability to handle a muddy track, are thought to be passed genetically from the parent animals to the offspring, the sire is often an important consideration in judging a horse who has not run enough to be judged on his own ability.

For our example: Sire - Dunkin Doughnut

Dam - This is the female parent of the horse, usually considered to also pass abilities to her offspring. There is a long standing debate among horsemen about whether the sire or dam have the most influence in contributing to the quality of the offspring. In the end, however, most handicappers agree that the horse's past performances are much more important than any royal lineage.

For our example: Dam - Sandy's Dame

Grandsire - This is the male parent of the dam, also believed to have a bearing on the characteristics of the horse in question.

For our example: Grandsire - Sandy Beach

Next comes information about the people who affect the horse and the way it is treated and raced: the breeder, the owner, and the trainer.

Breeder - This is located under the breeding line (sire - dam by grandsire). The individual who bred the horse is considered important by many, for some farms have a history of turning out quality horses almost all of the time. Included with the breeder is the state in which the horse was bred. (Some breeders run more than one farm, in different states. This is important, because many races are often restricted to horses bred in a certain state.)

For our example: Breeder - Sandy's Stables

Owner - This is the individual, or stable, that owns the horse. Also important as some stables have a knack for buying quality stock, and are consistently very successful at the track.

For our example: Owner - Sandy's Stables

Trainer - The person in charge of training the horse, and keeping it as close to perfect racing condition as possible. (The coach) One trainer may work for many owners, or one owner may employ different trainers for the different horses he owns. Trainers, like coaches, have styles, and a correct analysis of a trainer's style could indicate when he will have the horse in condition to win a race!

For our example: Trainer - Sandy O'Malley

Now we have a short summary of how successful the horse has been in it's racing endeavors. These "earning" records include the: Starts, wins, places, shows, and amount earned for different periods of the horse's racing career.

Earnings - This is the jumble of numbers located in the upper right-hand corner of the horse entry. What this basically is is the overall summary of how the horse has done in his races. What it includes are the numbers of starts, wins, places, and shows, and the amount of money earned for each of the following: lifetime (total), this year, last year, and turf. (Turf races are considered separately because horse's rarely can run equally well on both dirt and turf.)

For our example: Lifetime: 21 8 1 3 21,054 means:

21 starts, 8 wins, 1 places, 3 shows, and $21,054 dollars earned (total) in the horses entire racing career Similarly, for the current year (1985), the record is

1985: 4 2 0 1 5,460

and for the previous year (1984):

1984: 10 6 1 1 10,332

and for races on turf (total):

Turf: 1 0 0 0 0

Underneath the information describing the horse, come the horse's past performances lines. These are detailed descriptions of the "play by play" action and results of the horse's last eight races. The most recent race is listed at the top, as indicated by the respective dates on each line. For our examples, I'll use the latest race from the past performances for our sample horse "Munchkin".

The first group of information on a past performance line gives the unique identifying information about the race. This includes: the date the race was run, the race number, and the track.

Date - The date the race was run.
For our example: Date - 17Feb85

Race number - That's the number right after the dash (-). Most tracks have ten races on any given day.
For our example: Race number - 5 (fifth race of the day)

Track - This is the abbreviation for the track name where the race was one. "Suf" expands to really be "Suffolk Downs", which is located here in Boston, Mass. "Rkm", or "Rockingham Park" is the next closest track, located in Salem, N.H. "Bel", or "Belmont" is a track in New York.
For our example: Track - Bel (Belmont)

Next comes some of the more "physical" information about the race; the: track condition, the distance, and the track type, and the fractional times in which the race was run.

Track condition - This is the condition of the track when the race was run. The abbreviations are, for dirt tracks (from best to worst): "fst" - fast, "gd" - good, "sly" - sloppy, "my" - muddy, "sl" - slow, and "hy" - heavy. A frozen track is either described as "fr" - frozen, or "icy" -icy. For turf tracks "hd" - hard, "fm" - firm, "gd" - good, "sf" -soft, and "yi" - yielding.
For our example: Track condition - fst (fast)

Distance - How long the race was, usually anywhere from five furlongs, to two miles. If the symbol "f" follows the number, that indicates furlongs. Races at distances less than one mile are called "sprints", while races one mile or over are called "routes". Distances of 1 mile followed by tiny subscripts, either 40 or 70 indicate 1 mile and 40 or 70 yards. (Ex. 1^{40} - 1 mile, 40 yards)

For our example: Distance - 6f (6 furlongs, or 3/4 of a mile)

Track type - Most tracks only have one racing strip of dirt; however for the benefit of those tracks with many racing surfaces, a space is left for the track type, should it be other than just the normal dirt. Possibilities for this are: - inner dirt (for tracks with two dirt ovals), - outer dirt, - inner turf, and - outer turf.

For our example: Track type - (inner dirt)

Fractional times - These are the times of the leading horse at the first and second call, and the end of the race. For sprints, the first call is at the quarter mile (2f), and the second call at the half mile (4f). For routes, the first call is at the half mile, and the second call is at three-quarters of a mile (6f).

For our example:

First call - 22. (22 and 1/5 seconds at the quarter)
Second call - 46. (46 and 3/5 seconds at the half)
Final call - 1:12 (1 minute, 12 seconds even at the finish)

Following these "physical" information pieces are the more "administrative" facts about the race. These include: the restrictions, and the race type and class.

Restrictions - These are symbols which indicate what the entry restrictions were for the race. Possibilities are: - Fillies only (or more accurately, females only, since mares can enter these), - State-bred only (for horses who were all born in the specified state),
- otherwise restricted in some sense (not seen too often), 3 - Three and up (for horses ages three and up, only).

For our example: Restrictions - (Race for fillies, three and up)

Race type and class - There are different official types of races in which a horse may be entered. Some of these are as follows:

Clm - Claiming In a claiming race, all of the horses may be bought (claiming) before the race begins, for a specified class price. For example, in a "Clm 4000" race, horses can be bought for 4000 dollars. (If the owner is willing to sell for less, the horse will be allowed to carry a few pounds less in the race, to compensate.) What this type of race achieves is fairness - an owner will not enter a good horse in a cheap claiming race (where it's sure to win) because it may be claimed, and he would lose it. Thus horses are entered at the price their abilities are best suited for, and races are competitive. An owner will move a horse up or down in class, when the horse's performances indicate stronger or weaker competition is required.

Alw - Allowance A non-claiming race in which entries are regulated by previous purse earnings and/or number of victories. Allowance races come in stipulated classes, too, such as "Alw 8000".

Md - Maiden A race in which none of the horses have ever won a race. Maiden races indicated in this way, ex. "Md 4000" are for maidens, who also may be claimed (for 4,000 dollars)

Md Sp Wt - Maiden Special Weight A race for better maidens, where all of the horses carry the same weight (except for sex allowances, for fillies against colts).

Hcp - Handicap A special race, for some purse, of better horses. The Kentucky Derby would be an example of a handicap race. Officially, a handicap is a race in which the racing secretary or track handicapper assigns the horses different weights in an attempt to equalize the field.

When a small *c-* appears before the class amount it indicates that the horse was actually claimed before that race (somebody bought it).

For our example: Race type and class - Clm c-13000 (A claiming race for $13,000. The horse was claimed)

Now comes the play by play position calls for our horse (Munchkin), for different stages in the race. This include: the post position, and the position calls.

Post position - This is the first number of the series of numbers which follow the race type and class. This is the starting gate from which the horse started the race. Post position 1 is closest to the inside fence, and hence a great advantage over, say, post position 12. In some races, between otherwise equal contestants, the difference in post positions is the deciding factor.

For our example: Post position - 5 (fifth horse from the inside rail)

Position calls - These are the indicators of the horse's relative position during the race. The first number is the start call, which occurs a few yards after the starting gate. Next is the first position call and the respective lengths from the leader at that time. If the first call were, say, 5 , the horse would be 5th, and 4.5 lengths behind the leader. A length is the length of one horse, about 8 or 9 feet. When the horse you're studying IS the leader, the lengths are the distance from him to the first horse behind him. If the distance is smaller than a length, fractions such as , , and are used, along with

The position calls include the start, first call, second call, third call, and finish. In sprints, the start call occurs right after the starting gate, the first call at the quarter mile, the second call at the half mile, the third call at the stretch (1 furlong from the finish), and the finish call at the official end of the race. In routes, the start call is at the quarter mile, the first call at the half mile, the second call at three-quarters of a mile, and third call at the stretch, and the finish at the finish.

For our example: Position calls - 1, 1 by 1.5, 3 by a nose, 3 by .75, 3 by .75, indicates that:

Start - 1 (the horse broke first)
First call - 1 (first at the first-call, leading by 1.5 lengths)
Second call - 3 (third, and a nose behind the leader)
Third call - 3 (third, and¾lengths behind the leader)
Finish call - 3 (third, and¾lengths behind the leader)

The next group of information describes those facts which were available to the racing fans *before* the race was run. This is the information which handicappers on that day had to go by: the jockey, the equipment, the weight, and the odds.

Jockey - The jockey who rode the horse in that race.
For our example: O'Malley S.

Equipment - If the horse was wearing any extra equipment during the race, it is noted here. Possibilities are: "s" - spurs, and "b" - blinkers. (Blinkers are used to keep the horse's mind on racing, and also to not allow him to see things on the side, such as another horse, which may frighten him.)
For our example: Equipment - b (blinkers)

Weight - The weight the horse was required to carry in the race. Horse's are required to carry certain weights to even out their abilities with the rest of the field, and to make the race more competitive. Jockeys usually weigh less than the required weight, so the difference is made up with lead weights, which are placed in special pouches in the saddle. The jockey and saddle are weighted both before the race and afterwards; until this is done, the race results are not official.
For our example: Weight - 115 (pounds)

Odds - The next number is the horse's win odds, given in dollars and cents. This is the dollar amount that would be paid for each dollar bet on a winning ticket, if the horse won the race. An asterisk in front of the amount indicates that the horse was the favorite in the race. (NOTE: In addition to the money won on a winning bet, the amount of bet is also returned. example. A dollar bet, with .50 odds, would return 1.50 if the horse wins.)
*For our example: Odds - *1.60 (the horse was the favorite, and would pay back 1.60 for each dollar bet if it won.)*

The next two pieces of information try to indicate how well the horse ran, in absolute terms (compared to horses horses racing this distance on this track.) These are: the speed rating, and the track variant.

Speed Rating - This is a number assigned to the horse for this race, based on the time of the winner, and on how many lengths this horse was trailing. The time in which the track record at that distance was made is equal to 100 points. From this, 1 point is subtracted for each 1/5 of a second slower that the winner ran the race, and 1 point for every length that this horse was behind the leader.

For our example: Speed Rating - 84 (3 and 1/5 seconds slower than track record

Track Variant - The speed rating alone is not enough to judge a horse's performance by, as the track on the day of the race may be sloppy or slow, and the horse would have to be a superstar to get anywhere near the track record. For this reason, a track variant is included, which is a number calculated from the average of the difference from 100 of of the speed ratings of all winners for that day. A higher track variant implies a bad track, where all of the horses that day did poorly compared to the track records. Of course, it could also mean that mostly cheap, low-quality horses were running that day. If that were the case, a horse which had run in a high-quality race that day, would have an enormous track variant on it's record, along with it's speed rating; the result being that the horse looks better on paper than it really is. The track was not as bad as the track variant implied.

For our example: Track Variant - 19 (a moderately fast day)

Finally, we have the more personal information about the race. This includes the horses which came in first, second, and third, a comment about our horse's performance, and the number of horses in the race.

Horses in the money - These are the listings of the horses which came in first, second, and third in that race, along with their respective weights, and lengths. In this listing, all of the lengths are distances to the following horse! For the first horse, the lengths would be the distance between it and the second horse. For the second horse, the lengths would be the distance between it and the third horse. etc.

For our example: Horses in the money:
Jim Dandy110 ChurchChk110 Munchkin115
First horse: Jim Dandy, carrying 110, by lengths
Second horse: Church Chk, carrying 110, by lengths
Third horse: Munchkin, carrying 115, by lengths

Comment - This is an expert's comment on how the horse really performed in the race. For example, "Bid but hung" means the horse made a good effort early in the race, but didn't finish it. On the other hand, "Gamely" indicates a solid try throughout the race, while "Driving" means a strenuous victory, under heavy punishment (by the jockey).

For our example: Comment - Early foot (quick start, but faded)

Number of horses - The number of horses in the race.

For our example: Number of horses - 7

In addition to the past performances, the horse may also have done some recent workouts. These workouts are also included with the horse entry, after the past performances. They are useful in some small extent to determining a horse's current condition, especially after a layoff (more than thirty one days without a race). Each workout line contains the following information. For our example, use the latest workout for Munchkin (located on the line under the past performances).

Date - The date of the workout.
For our example: Date - 5Mar85

Track - The track where the workout took place.
For our example: Track - Bel (Belmont)

Track type - The track type, as describe above.
For our example: Track type - (not described, which implies dirt)

Distance - The distance of the workout. Usually an integer number of furlongs, from 2 to 7.
For our example: Distance - 4f (4 furlongs, or half a mile)

Track condition - The condition of the track the day of the workout.
For our example: Track condition - fst (fast)

Time - The time for the workout.
For our example: Time - 50 (50 and 2/5 seconds)

Workout Style - Some times, in a workout, the horse is not running a fast as it can. Workouts range from easy to strenuous. The observing expert characterizes these styles as one of:

"b" - breezing (light, no encouragement)
"h" - handily (medium, hand encouragement, no whipping)
"d" - driving (hard, heavy encouragement, with whipping)
"e" - easily (mild encouragement)

The symbol "g" may also appear, indicating that the horse started from a gate. The symbol "trt" following the track abbreviation, indicates the horse ran on a training track. "tc" indicates turf course.
For our example: Workout style - bg (breezing, using a starting gate)

The Pari-mutuel Betting System

Now that we can read a horse entry and the corresponding past performances intelligently, or at least comprehensibly, the next thing on everyone's mind is how does one make money by finding the best horse. This is done by betting on it, and the betting system used at most major and minor tracks around the world today is the "pari-mutuel betting" system. Before rushing into a discussion of the betting system I need to define a couple terms I'll be using.

Pari-mutuel betting - a system of betting where all of the winners get the money bet by the losers, minus the track take, including taxes.

Odds - a figure consisting of two numbers, x and y, where the odds x - y (read as "x to y"), means x dollars profit on y dollars bet. A horse having win odds of 5 - 2 will return a profit of $5 for every $2 bet, if it wins.

Pool - the amount of money bet by all the bettors. Types of pools are: win pool, place pool, show pool.

Track take - the amount of money removed from the pool before it is divided among the winners. The "take" is usually 18%, some of which goes to the track, and the rest going to the state government.

Dime-breakage - the procedure of rounding down, to the nearest dime, the profit per dollar paid back to the bettor. While dime-breakage is common throughout the United States, nickel-breakage is the standard in Canada.

Paid - When used in the context of "the horse paid x", this means that x is the total of $2 and the profit made on a $2 bet. As $2 is the minimum bet at most thoroughbred tracks, it is the standard for describing the horse's return at the window. For example, if the profit made on a $2 bet was $4, then the horse *paid* $6.

In pari-mutuel betting, the odds on a horse are proportional to how many dollars are actually bet on the horse, relative to the other horses in the race. Let us look at a simple, but not totally accurate example, a two horse race involving Speedy Sam, and Slow Sid. This example is not totally accurate as we are neglecting such important considerations as taxes and dime-breakage.

Suppose all the bettors together had bet $200 on Speedy Sam to win, and only $100 dollars on Slow Sid to win. The win pool would thus contain $300. Let us now calculate the odds on Speedy Sam. First, assuming that Speedy Sam won, all the bettors who bet on him would get their money back. This is a requirement! Thus the total amount of money left to comprise the bettor's winnings is the pool minus the amount bet on the winning horse, Speedy Sam, which is now $100. This amount of money is divided among the 200 dollars bet on the horse, which comes out to be 50 cents per $1 bet. The bettor who had bet $5 on Speedy Sam would receive back his initial 5 dollars, and then 5 times 50 cents, or $2.50 profit. Thus on a $5 bet, Speedy Sam returned $7.50 at the window. Or speaking in more standard terms, Speedy Sam *paid* $3, (which is, as defined above, is the sum of the $1 profit made on a $2 bet, and original $2).

The odds for the two horses in the race would be as follows. For Speedy Sam, who earned 50 cents per dollars, the odds would be 1 - 2, The prospective bettor understands from these odds that Speedy Sam would return a profit $1 for every $2 dollars bet (or 50 cents for every $1). Our other champion horse, Slow Sid, did not seem to impress the bettors very much. If Slow Sid won, a bettor who bet on him to win would receive for every $1 bet: 1) the original $1 back, and 2) a profit of $2 ((the pool of $300 - $100 original bets) / $100 dollars bet). Thus Slow Sid would pay $6, and the odds for him would be 2 - 1.

That's the way it would work, if tracks made all their profit from admission prices, and governments didn't ask for their share. Unfortunately that's not the way things happen. Before the pool is divided, the track take must be removed. Referring back to our example, from our $300 pool, 18% automatically disappears, leaving the bettors with a "real" pool of $246. Now, supposing that Speedy Sam wins, the bettors get back their $200 dollars, and only $46 dollars is left to divide as profit. This turns out to be 23 cents per dollar. Not good, but it gets even worse.

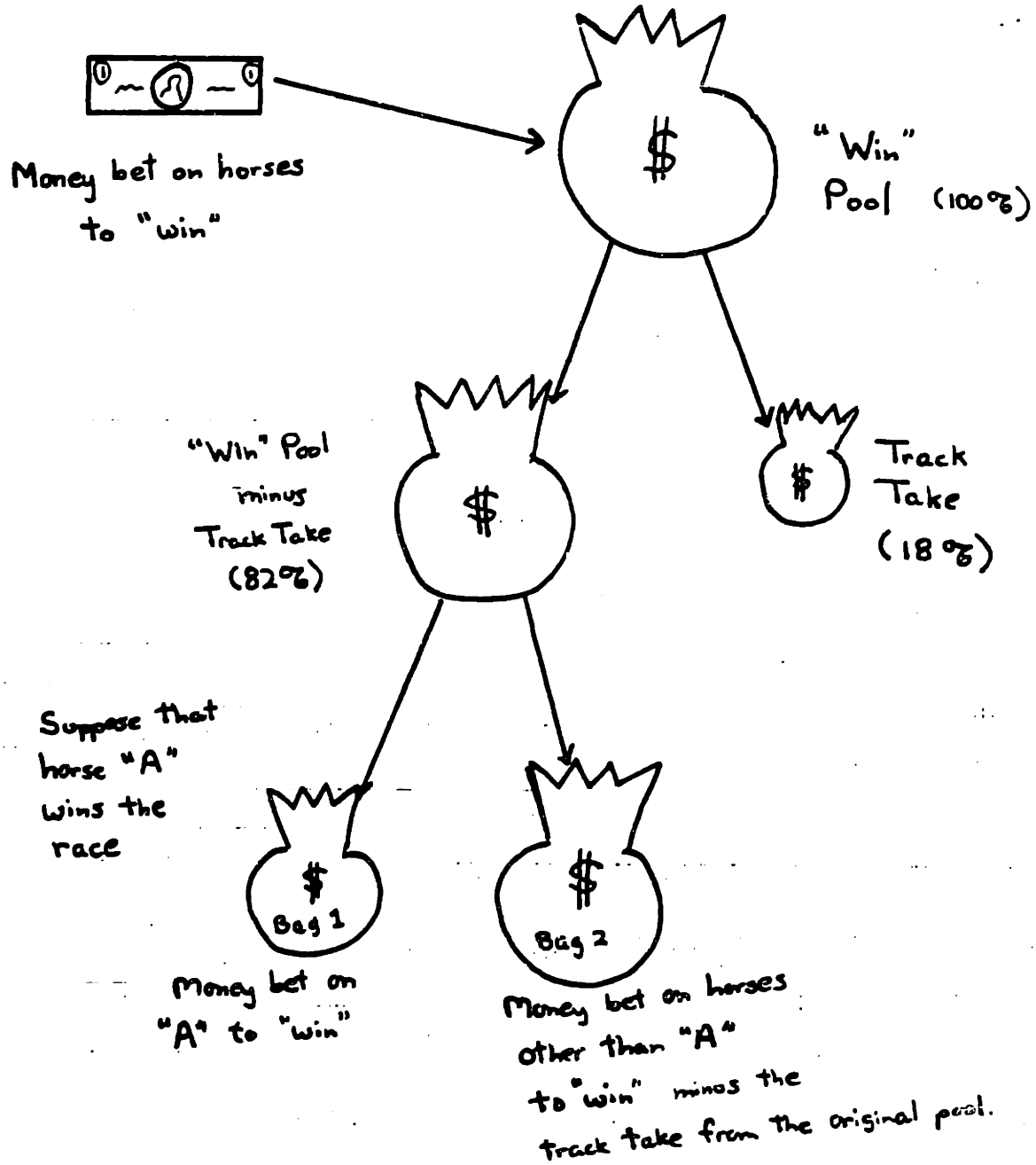
The employees who run the betting windows would rather not be bothered returning pennies or nickels to bettors, thus the profits per dollar are subject to dime-breakage. If Speedy Sam wins, he really only returns 20 cents per dollar bet, "pays" $2.40, and has odds of 1 - 5 ($1 profit for every $5 bet). This is in contrast to the original 50 cents per $1, which was how much Speedy Sam would have returned, had not the government and the track taken their share.

Slow Sid fares a little better, after track take and dime-breakage. The "real"pool of $246 returns $100 back to the bettors, and then divides $146 over the $100 dollars bet. The profit on Slow Sid is therefore $1.40 per $1 bet (dime-breakage on what should be $1.46); he pays $4.80, and has odds of 7 - 5.

If this is still a bit confusing, please consult Figure 2 for a digram depicting how the original win pool is subdivided and diminished until it finally returns back to the bettors.

That's as bad as it gets, folks, and it probably seems now that there really isn't much reason to study handicapping, as the profits, if there are any, are so low. This is true, for a two horse race. This is why most races have six to twelve horses! In a six-horse race, the money bet on the five losers becomes available, after taxes and breakage, to be divided among those who bet on the winner. And if the winner was not the favorite (the horse with the most money bet on him), the pay-offs could be quite substantial. Do not despair! One professional handicapper, Andy Beyer, who also

Figure 2: Distribution of the Win pool



Money Profit per dollar bet = $\frac{\text{Bag 2}}{\text{Bag 1}}$ − Dime Breakage

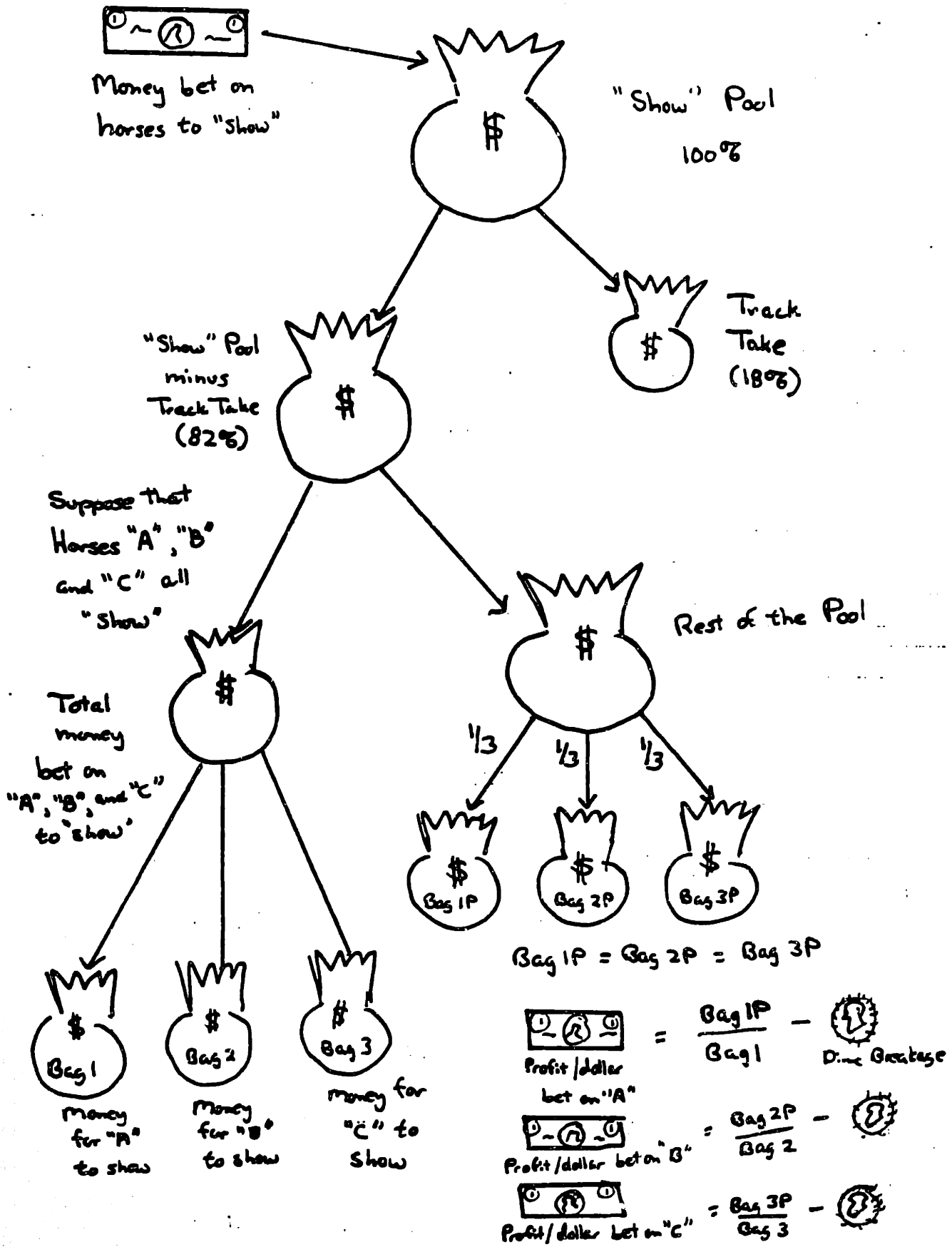
works for the *Daily Racing Form*, made a profit of over $50,000 in 1977. [Beyr78]

But wait, there's more! In a twelve horse field, or even a six horse field, it sometimes gets tricky to find the one best horse. Hence the advent of what is know as "place" and "show" betting. In races with four horses or more, place betting is also offered, and in races of 6 or more horses, show betting is offered, in addition to place betting.

If a horse is bet for place, the bettor collects if the horse comes in either first or second. A horse bet for show returns money if the horse comes in either first, second, or third. Separate pools are maintained for win, place, and show bets. The pools are divided in the following manner. First and foremost, off comes the track take. Then from what is left, the bettors who bet the first horse "to place", and the second horse "to place" are returned their original bets. Whatever is left is divided in half equally to be distributed as profit to those people who bet place on the horses that came in first and second. An example is worth a thousand words, or at least a couple hundred.

Two of the horses in an eight horse race were Steppin Sue, and Grace Swift. At the wire, Steppin Sue won by a nose, with Grace Swift right behind. While Steppin Sue won, both of the horses placed. Now suppose the total place pool was $15,000 before taxes, with $3,000 bet on Steppin Sue to place, and $1,000 Grace Swift to place. The pool, after track take, would be $12,300. First the original money bet is taken out, thus leaving the pool at $8,300. This is divided in half, leaving $4150 dollars to the bettors of each horse. Those who bet on Steppin Sue, all $3,000 dollars worth of them, each get ($4150 / $3000, with dime-breakage) or $1.30 per dollar bet to place. Thus Steppin Sue paid $4.60 to place. Only $1,000 was bet on Grace Swift to place, which makes her a bargain, now that she did come in second. Each bettor who bet her to place with receive ($4150 / $1000 with dime-breakage), or $4.10 per dollar bet. Thus Grace Swift paid $10.20 to place. Not bad, $8.20 profit on a two dollar bet!

Figure 3: Distribution of the Show pool



Show betting works in the same way, except that there are three horses which show, and the pool is divided in three instead of two. See Figure 3 for a diagram of how the money flows from a "show" pool. As would be expected, a horse that wins should pay the most when bet to win, a lesser amount on a bet to place, and even less when bet to show. This is not always the case, however, and at some of the races I've witnessed out at Suffolk Downs, winning horses have paid more to place than to win. This usually happens when an uninformed betting public automatically bets the favorite to win, and the next best horse to place, not realizing, or forgetting, that if the favorite wins, he will also place and show. Hence, there is often compensation for the bettor who watches the pools carefully and determines the place and show odds for horses, as well as the win odds. (At tracks, the winning odds are displayed, as well as the pool totals and the amount bet on each horse, for each of the win, place, and show pools. Why the place and show odds were not also displayed was a source of puzzlement to me for a long time, until one day I realized that the odds could not be calculated until after the race. One did not know which two horses would be coming in for place, or which three for show, and therefore the amount of money which would have to be returned to the bettors for the placing (or showing) horses could not be subtracted from the pools (a step necessary for determining the odds) could not be determined until one knew which other horse had also placed, or which other two horses had showed.)

Other types of betting pools exist, however I will not go into them, save for describing their general makeup. For Quinella betting, the bettor picks two horses, which must come in first and second, but in either order. In Exacta (sometimes called Perfecta) wagering, the bettor must, on his bet, pick the first and second horses, in order. If the two horses come in, but in the wrong order, his bet does not win. For a Trifecta, the bettor picks the first, second, and third horses, and order counts. For a Daily Double, the bettor picks the first horses of two different races. For a Twin Trifecta, the bettor must pick the first, second, and third horses, in order, for two

separate races (six horses total)! Payoffs for these "exotic" types of betting are often very high, sometimes in the thousands of dollars for trifectas, however the bettor must be cautious and very certain of his selections, for the odds of mishap are always high in horse races, and the chances of a mishap occurring, such as a stumble, or a bad start, are much higher when spread over three horses. Caveat emptor!!

2. Developments in Analytical Hnadicapping

This chapter describes some of the developments in handicapping over the last twenty years and then indicates how my computer analysis program will advance the state of the art of handicapping.

Although thoroughbred racing reached it's peak period of expansion shortly after World War II, organized volumes of information about the sport were not readily available to the aspiring handicapper. Granted, there were a great number of pamphlets which were suppose to reveal the "secrets" of handicapping. These usually turned out to be either hoaxes or limited discussions about various handicapping angles. None of them resembled anything like a comprehensive study of all of the factors which influenced the outcome of a horse race. As one merchant explained it, "The reason they don't publish books for horseplayers, is that horseplayers can't read." [Ains68]

The first real effort at producing a complete book which explained a handicapping method and the reasons behind it was *The Compleat Horseplayer*, written by Tom Ainslie in 1966. This book was complemented by *Ainslie's Jockey Book*, written in 1967, which explained another version of the same method of handicapping, with more emphasis on the importance of the jockey. [Ains68] While both books quickly became best-sellers among handicappers, they did not teach a handicapper how to understand thoroughly the factors involved, and how they related to one another. The books taught handicappers a method of handicapping; however, they did not teach them **how** to handicap! What handicappers needed was not somebody else's method, but the tools to develop their own systems.

This need was recognized once again by Tom Ainslie who then wrote the handicapping masterpiece *Ainslie's Complete Guide to Thoroughbred Racing* in 1968. Instead of just producing a handicapping method, and explaining how it worked, this handicapping encyclopedia thoroughly discussed the physical layout of racetracks, the arithmetics of the pari-mutuel system, the ways in which breeding affects the

configurations of the thoroughbred, and what a good racehorse should look like, and why. It's central section explored the arts of handicapping; schools of thought dealing with distance, form, class, age, sex, consistency, weight, speed, pace, post position, the owner, the trainer, the jockey, the paddock, the post-parade, track variants, speed handicapping, and many more were all analyzed in detail. The handicapper could now get some concrete idea on what to look for and how all of the factors fit together in analyzing a race. In addition to giving solid ideas of what to look for, Ainslie provided in his book seventy-seven spot methods for picking a winning horse. Each of these methods described some conditions to look for in a horse and the race. When all the conditions are true, the intelligent bettor should bet on the horse. [Ains68] These methods were all very good at picking out solid combinations of important handicapping factors which fit together in a somewhat predictable manner. Some of the problems with them, however, were 1) the obvious tedium at checking all of the conditions for each of the seventy-seven methods for each horse in a race, and 2) the resulting few races in which any of the methods proved to be applicable. Still, they were very useful rules of thumb, which a thorough handicapper would find very beneficial to aid him in his handicapping.

Ainslie's Complete Guide to Thoroughbred Racing remained the one and only reliable source for handicappers for much of the following decade. Then, in 1975, a Harvard-educated, professional handicapper, Andrew Beyer, published a horseplayer's guide *Picking Winners.* (An interesting note is that while Andrew Beyer attended Harvard, he did not graduate due to a conflict between his finals and the last race of the Thoroughbred Triple Crown!) *[4]* Beyer's largest contribution to the art of handicapping was the organization of the school of thought known as "Speed handicapping." Andrew Beyer described how, by recording the final times for every race at a track, every day, a horseplayer could comprise par tables which contained the average time at each distance for each quality of horse. Then, an accurate track variant

could be comprised from these figures, and from these a speed figure for the horse, which is a number which indicates how well the horse performed, speed-wise in a past race. By computing speed figures for all of the horses in a race, one can accurately determine which horse is really the fastest. Beyer was quick to point out that this does not always indicate the best horse, because such considerations as current form and jockey have a great bearing on how a horse will do in today's race. However, it does give a solid starting point for comparing horses. Another contribution of *Picking Winners* is a discussion of track biases, the condition where one section of a racing track is noticeably faster or slower that the other sections. Accurate predictions of which horse will get stuck on that section of track is another way of picking winners, or eliminating possible contenders. [Beyr75]

Another relatively complete volume discussing handicapping issues was *Betting Thoroughbreds* published in 1977 by Steven Davidowitz, a contemporary, and friend of Andrew Beyer. While *Betting Thoroughbreds* mostly rehashed the factors involved in analyzing a race, he did introduce two new concepts to the state of the art. The first was the idea of the "trip" handicapping. The important issue was not how fast the horse ran the race, or what his speed figure was, but **how** the horse ran the race, i.e. what type of trip did it have. Was the horse trapped behind slower horses at some point in the race, or did it reach the turns on the rail, or five horses away from it? Did the horse show amazing burst of speed in attempts to get out of trouble during the race, or did it have to pull up sharply to stay out of trouble, and to go around other horses? A horse which had a bad trip, but still did reasonably good was sure to have better luck in it's next race, thus providing a good bet. The second new idea was similar to trip handicapping, but with the added luxury of not having to physically observe each race (a necessary step for trip handicapping). This was the idea of a "key" race; a race in which the horse did something unusual for it to do in either one of it's latest workouts, or in it's last race, and then came back to win today's race. For example, a horse which normally had slow

starts in all of it's previous races suddenly had a fast start in the race previous to a wining race. The method of doing this is to take each winner of a race, and then search back to it's last race or workout, and see how the horse ran the race. If enough winners showed some consistent factor in the race previous to the winning one, a handicapper may to on to a handicapping tool which isolates a type of well-run which occurs previous to a winning race. [Davd77]

Another notable book in the evolution of the art of handicapping was *My $50,000 Year at the Races*, published in 1978 by Andrew Beyer. This book highlighted Beyer's 1977 season at the tracks, in which he netted a $50,000 profit, thus proving that "Maybe you can't beat a race, but you can beat the races!" While this book did not introduce any new insights to the game, it did provide numerous valuable examples on how to analyze races using speed handicapping, track biases, and trip handicapping. [Beyr78]

Finally, in 1979, the power of the computer was first applied in an organized manner to the problem of thoroughbred handicapping. Dr. William Quirin, Ph.D., and Associate Professor of Mathematics (Computer Sciences) at Adelphi University conducted a computer analysis of the results of the races of five thousand horse. His results were published in *Winning at the Races: Computer Discoveries in Thoroughbred Handicapping.* What Dr. Quirin did was to take every popular handicapping theory and angle, and see what percentage of horses with that characteristic actually won their next race. In addition to the percentage, he also provides the dollars net, which a $2 bet on every horse with that characteristic would have yielded. This very organized study of every major handicapping angle enabled handicappers to sort out those angles which were both most effective, and also which returned the highest profit. In addition to the charts analyzing each factor, Dr. Quirin also provided two "computer-generated" regression formulae, one for sprints, and one for routes. The formulae employed those handicapping angles found to be most successful in the races which the computer studied. [Qurn79]

Another notable volume which made a breakthrough in thoroughbred handicapping is *The Body Language of Horses*, published in 1980, by Tom Ainslie and horse trainer Bonnie Ledbetter. While other handicapping guides had briefly discussed what to look for in a horse at the track, this was the first complete, comprehensive guide written for the novice and the professional handicapper, alike, on just what specifics to look for in horses at the track. This guide discusses equine behavior in all the stages of the pre-race formalities; positive and negative factors in everything from coming out of the barn, to saddling up, to the post parade, to warming up, are described in detail. In addition, advice is given on how to spot a horse which has been given either barbiturates, or amphetamines. [Ains80] This reference guide provided invaluable advice to the handicapper who frequently became frustrated when the horse that looked the best on paper, failed miserably. Now there was a way to predict, by looking at a horse, what his condition was like today, and how he might run the race!

One last major work which contributed to the science of handicapping was *The Winning Horseplayer*, written,once again, by Andrew Beyer in 1983. In this book, Beyer described, more thoroughly, how to become a proficient trip handicapper, and then how to relate trips to speed figures in an attempt to improve overall handicapping efforts. He also discussed a topic which had remained taboo for many years: How a horseplayers mental attitude affects his handicapping efforts. Beyer had found through his own experiences, and those of his contemporaries, that a handicapper's mental attitude greatly affects his style of handicapping races, and the corresponding success of the handicapping attempts. Horseplayers who are currently on winning streaks often remain so, handicapping races with such precision and insight, that they often amaze themselves. The opposite is also true, however, and losing streaks also tend to perpetuate themselves. Beyer concludes in *The Winning Horseplayer* with words of advice about how a handicapper should try to avoid these dark spells by recognizing their existence, and the corresponding decrease in handicapping skill. [Beyr83]

Now that I have discussed how the art of handicapping has evolved from the confusion of the dark ages, to the well-defined art that it is today, I would like to divulge my own sentiments about thoroughbred handicapping. I believe that a computer could be turned into a very powerful handicapping tool, which could analyze races much more thoroughly than handicappers could by hand, and whose expertise was only limited by which factors and angles it had been programmed to analyze.

Dr. Quirin's formulae are the first results of a computer analysis of thoroughbred handicapping. However, they did not prove to be an end unto themselves. Handicapping with them proved to be very slow and tedious, often taking three to five hours to analyze one race. In addition, they did not take into account various effective "spot plays", and rules of thumb, proven effective by Tom Ainslie back in 1966, and later in 1980, when he revised and updated *Ainslie's Complete Guide to Thoroughbred Racing.* These spot plays, which occurred successfully, but infrequently, were averaged out of Quirin's computer-generated formulae. The unfortunate result of all of this, is that a handicapper following Quirin's formulae will overlook some of the outstanding longshots produced by Ainslie's rules of thumb. I believe that better results could be had if the two handicapping methods were combined to produce a system that looked for the infrequent conditions that produced good longshots, and, not finding these, proceeded to handicap the race in as best a way as possible, using Quirin's formulae.

The advantages of using a computer to actually handicap races are numerous. First of all, the time necessary to apply Quirin's formulae in handicapping a race is on the average, several hours. Now, if one were to also apply each of Tom Ainslie's seventy-seven rules of thumb to the race, the time quickly blows up to anywhere from five to ten hours per race (after sufficient practice with both systems to get up to this speed). Then of course, both systems require many detailed calculations, any of which, when done incorrectly, could cause a handicapper to lose his shirt. A computer could be used to handicap a race in time considerably faster than by hand, and without the

possibility of human errors.

Secondly, there is the question of consistency. This was the problem pointed out by Andrew Beyer, where a handicapper's mental attitude greatly affects his handicapping skills. [Beyr83] A computer would have no such problem, and would be especially consistent, and equally picky about details in every race it handicapped.

Next, there are the various other handicapping methods suggested by the professionals, such as searching for "key" races, or trainer patterns, which could be done with great ease using a computer system with a reasonable data base. Although these handicapping methods are some of the best, they are never suggested for anyone other than the full-time, professional handicapper, due to the time involved in finding the data, and trying to develop patterns. This would not be a problem, using a computer as an analysis tool, provided that our computer system set up its data base in such a manner that searching for matches between horses and a trainer, or a key race (or any piece of horse data), could be accomplished quickly and easily, with minimal interactive effort from the user.

A fourth advantage of a computer handicapping tool, is the possibility to, with relative ease, apply several handicapping methods to a race, and determine which are most successful. Perhaps the success of any method is dependent on characteristics of the race. Would one system work better on low quality horses, while another be more successful on higher class horses? Maybe one system proves useful for maiden races, while another is more applicable for races with older horses. Or is the distance of the race the crucial factor, in deciding which analysis to perform on a race? Maybe we should just apply every analysis we can get our hands on to every race, in some weighted formula as to how much the results of each analysis count. Maybe the weight should depend on the conditions of the race, such as distance, and class. The possibilities are endless! With a computer handicapping tool, general enough to easily manipulate data and analyses, many of these possibilities could be explored, perhaps

finding correlations to one or more factors.

A fifth possibility would be just to load up every possible analysis into our computer, and have it analysis some large number of races. For each race, we would also provide the results, so that the computer could determine how close it had come to finding the correct solution. It could then go back to each analysis, and determine which tests, rules of thumbs, or methods had picked out the winning horse as being the best. In some fashion, it could be possible for the computer to determine the best weights for each of the handicapping methods, given the conditions of a race. And each time the computer made a mistake, it would update it's knowledge of the world in an attempt to create a better handicapping algorithm.

This type of "learning" program could be created, if the best model to describe how the factors relate to each other really is linear in nature. But suppose that it isn't, as well may be the case. Our computer system, to be truly general in nature, must not be limited to just to one model of a horserace analysis. Ideally, we want our handicapping tool to be able to explore other possible models, as chosen by the user. What we are really trying to create is a flexible, interactive environment, in which the user has the ability to easily create models for thinking about horseraces, and then implement them, using the working environment which we have established. For example, maybe one such model would involve the exploration of one factor, and then further explorations of different factors, where the method of exploration would depend on the results of previous tests. Our system must provide the user easy access to all of the factors he may need for developing new models for analysis, and the variability to use them in what every manner he chooses. In this way, our computer system really would become a useful tool for the development of models for horseraces, and for the implementation of these models.

Thus, the idea for **WHINNY** was born! **WHINNY** would be a computer program with a friendly user interface for the entry of horse data and past performances, the ability to put together races using horses for which it had information, and then to handicap races, using one or more analysis methods, either separately, or in some weighted formula. **WHINNY** would also be general enough to support the addition of new models for thinking about horseraces, and the easy implementation of these models. Such a computer program would be an invaluable handicapping tool, both for those handicappers who wish to quickly and easily handicap a race, and for those who would advance available handicapping techniques by either improving existing methods, or perhaps developing new ones.

WHINNY would not, however, be an end unto itself. My goal in this thesis is not to create the world's best handicapping program. While this would be nice, it is not possible yet, because the necessary tools for the entry of, and structure and management of the data relevant to the problem does not yet exist. **WHINNY** is to be that tool, the result of applying modern computer power and interactive software to produce an environment in which the necessary information for handicappers is readily available and manageable. With the aid of **WHINNY**, in the future handicapping strategies may be combined and compared, perhaps to the discovery of the "best" computer handicapping program. This thesis is a beginning. rather than a "last word" in this effort, and the optimization of betting strategies or handicapping strategies is a problem which will require further investigation by handicappers and other interested individuals in the field (or at the track).

3. WHINNY: Computer Handicapping Tool

After I had decided that a computer handicapping program would be useful, not only for analyzing thoroughbred races, but also for advancing and improving existing handicapping techniques, I had to decide how to design and implement the WHINNY system. How the system should be designed was not immediately apparent, as there were many ways the problem could initially be approached. The general structure became more and more clear, as I considered some of the relevant issues such as ease of data entry, and the ability to create many different methods for race analysis. Here are the features which I believed would be vital to the usefulness of the WHINNY system.

First, the interface for data entry would have to be easy to use, and as automated as possible, for there was going to be a lot of data to enter. For each horse entry, including past performances, there is on the order of four hundred individual pieces of data. A race could have anywhere from six to twelve horses in it. When one also considers the fact that most tracks have ten races per day, and have racing three to six days a week, the problem of data entry seems insurmountable! Well, maybe it was, but I wasn't going to give up without an honest try. Many of the data pieces were limited in nature, and could be narrowed down to one of several possibilities. My data entry system was going to have to make the most of this, by presenting the user with well-defined choices everywhere it was possible to predetermine the finite set of possible selections. In addition, if any piece of data could be calculated from previous data, or had the possibility of already existing in the database, the program would have to try to calculate or find it, before asking the user to enter it. Also, at any point in the data entry, the user must have the ability to go back and correct mistakes which he may have made in the entering of any data piece.

A second important issue is that the data must be structured in such a manner that the program may retrieve any piece of information about a horse by just having the horse's name. If information was needed from a horse's past race or workout, just the

horse name and the date of the past race, or workout should be necessary.

A third, and very important feature, is the generality of the handicapping analyses. For a truly versatile system, it should be possible to create analyses, which are, either in part or in whole, composed of similar analyses. One could then test various combinations of analyses, with some assigned weights, just by creating a new analysis which contains the ones you wish to test, with their corresponding weights. For this to be possible, all the handicapping analysis should receive and return data in a standard format.

Before actually proceeding to describe how I actually designed and implemented the system, this is probably a good place to quickly describe the computer system I chose to implement WHINNY on. For reasons which will soon become apparent, I decide to use a *Symbolics 3600* Lisp Machine, writing the program in ZetaLisp, and using a high-density bit-map screen for my display. For those readers who are unfamiliar with the *Symbolics 3600*, Never Fear! I fully intend to explain any of the features of the Lisp Machine which I used directly to support the implementation of my program. For those hackers with many hours logged on this sophisticated piece of machinery, Read Swiftly and Enjoy!

Data Entry using Mousable Menus

The first problem which I attacked was data entry. If no reasonable method for the quick and easy entry of all the necessary data, solving the other problems would be somewhat of a mute point. One point in my favor, was the fact that every horse I would wish to study would have the same data fields as any other horse. As mentioned previously, the possibilities for many of the data fields are very limited in nature. For example, there are a small number of acceptable choices for a horse's color, sex, and age. When entering past performances, items such as track name, track condition, track type, distance of race, race type, race class, and many others are also limited in choices.

If at the correct point in data entry, the program could present the user with the acceptable choices for a particular data slot, the user could merely select the correct choice, without the undo effort of typing the same information, over and over again.

Fortunately for me, such a method of selection from a finite list of choices was currently available, in the form of mousable item menus. Starting with basics, a mouse is a pointing device attached to some computers, with the ability of moving a cursor over the computer screen, when the mouse is rolled across a surface. A mouse also equipped with a number of buttons, usually three, which, when pressed, cause something to be done to or with the thing which the mouse is pointing at. As an example, the *Apple McIntosh* personal computer uses a mouse as a pointing device to refer to some item displayed on it's screen. Not all items displayed on the computer screen need be "mouse-sensitive". However, I will want my choices of data for each slot to be "mouse-sensitive", so that by displaying all of the possible choices on the screen, I can pick the one I want by pointing at it with the mouse and clicking one of the buttons. *[5]*

This is how I implemented my choices for the data slots. For each data slot, all of the possible choices are grouped together as mouse-sensitive items in a "menu". There are different menus for the data fields for horses, past performances, and workouts. The WHINNY input system allows the user to click on each menu by pointing at the correct piece of data, and pushing a mouse button to enter the data for each correct field of horse being entered.

One additional feature made this use of mousable item menus even easier. A rather clever hacker from the Artificial Intelligence Lab at M.I.T, put together some code, which would cause menus to have highlighted borders; that is, the outer edges of the menu windows would become darker or lighter, having the effect making that menu stand out amidst the others. This feature enabled my program to highlight the window from which the next data item should come, thus helping the user to select the correct

menu. If another menu was clicked upon instead, in an attempt by the user to enter incorrect data, the correct window would flash it's highlighted borders indignantly for a second, to direct the user back to the correct menu.

In addition to having permanent mousable item menus which remained in a fixed position on the screen, it is also possible to created temporary mousable item menus. These menus pop onto the screen at the current location of the mouse cursor, and exist until either the one of the items is clicked on, or the mouse cursor is moved off of the temporary window. At this point, they disappear from the screen. These temporary menus are very useful in providing menus which are too big to normally keep on the screen, or not used frequently enough to warrant keeping on the screen, or of a changing size, where it would be impossible to determine until the window is needed, how big it should be.

Mousable item menus even were useful in situations where a complete set of all possible choices did not exist. This was done by creating code which would either accept a mouse click from the correct menu, or characters typed on the keyboard as proper input. Checks would be done, of course, to determine that the data entered from the keyboard was at least of the expected type, i.e. either a character, or a number, or a string.

In summary, I used the *Symbolics 3600* because it allowed me to set up an interface in which the user merely has to point at the correct data piece, and click on a mouse button, to have that data entered into a horse entry. In the next section I will describe how I actually used this "mousable menu" feature to set up menus on my Lisp Machine screen from which to make my selections.

The WHINNY Frame

The *Symbolics 3600* has a very sophisticated window system, which allows you to divide the display screen into many different window, each having different properties, if necessary. I devised a *Whinny Frame*, a screen arrangement which divided the screen into four horizontal portions. See Figure 4. At the top of the screen, there was a display window, on which a horse and it's past performances and workouts could be displayed. Races could be displayed here, too, complete with horses entered in the race, their weights and their jockeys. Analysis could also be displayed here, showing the user which tests or analyses they employed, and by what percentages did they weight the tests or analyses. In the middle of the screen, I put all of the mousable item menus, from which I could choose data. Underneath this, was a command line, which really was a special mousable item menu consisting of all of the commands the program could be asked to do. At the bottom of the screen, I located an interactive keyboard window, on which characters could be typed to enter data directly through the keyboard. This window would also be used to display any messages the program might send the user, such as which field of data it was expecting, or what additional information the program needed.

Those menus which permanently located in the *Whinny Frame* are as follows. First, on the left of the menu section of the frame are three "data" menus, containing respectively, days, months, and years. These three may be clicked upon in any order, and are originally all highlighted. As each one is clicked on, it turns it's highlighting off, until all three pieces of the data are entered. Next to the data menus is a "Race Number" menu, which contains the numbers (1 - 12). To the right of this is the "Track" menu, which contains abbreviations for some of the more common tracks which I would be entering. At the bottom of this menu is the mousable item *Other*, which, when clicked on produces a temporary mousable item menu containing all possible U.S. tracks. For an example of what a temporary mousable item menu looks

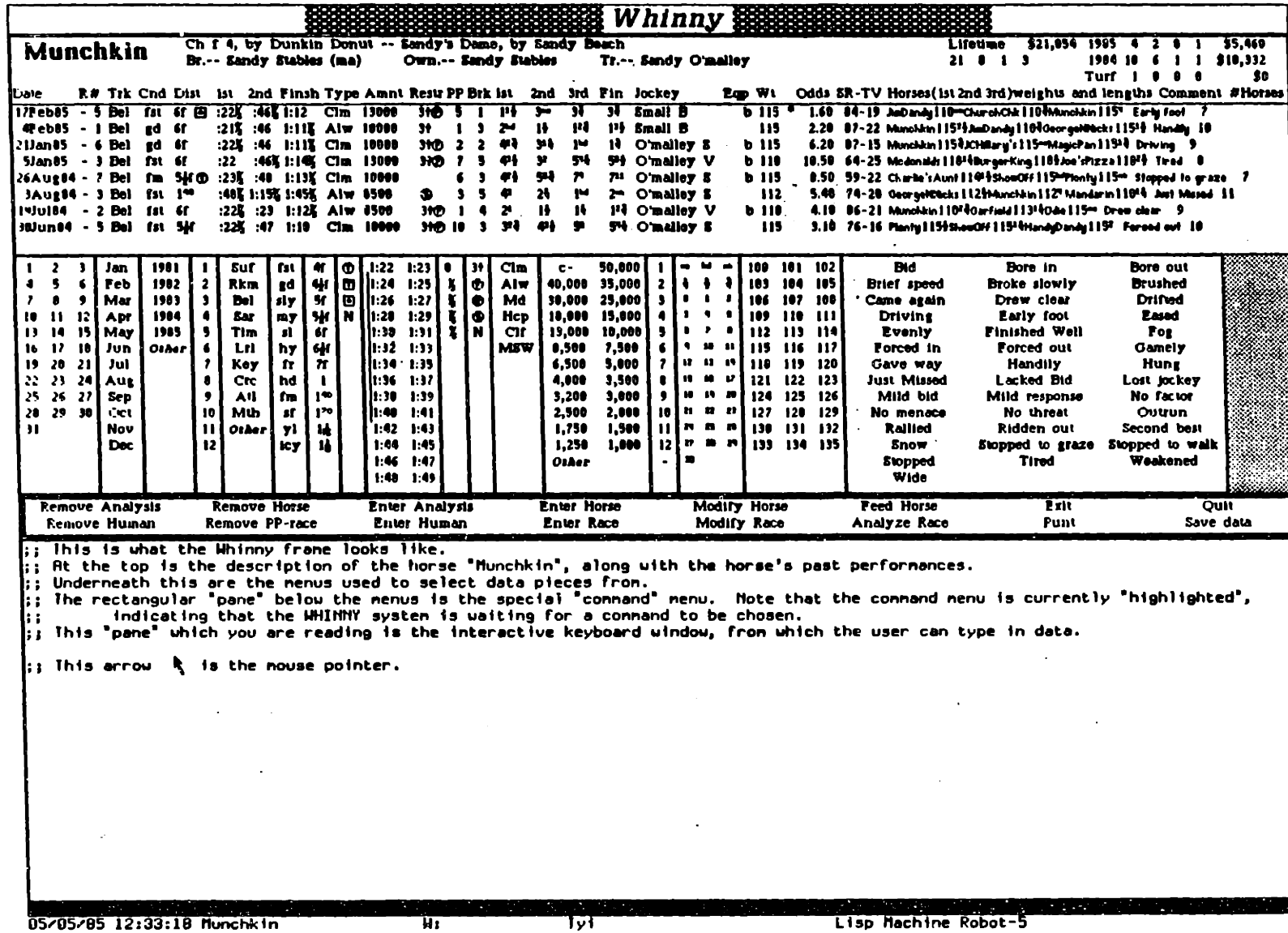


Figure 4: *Whinny frame displaying a horse*

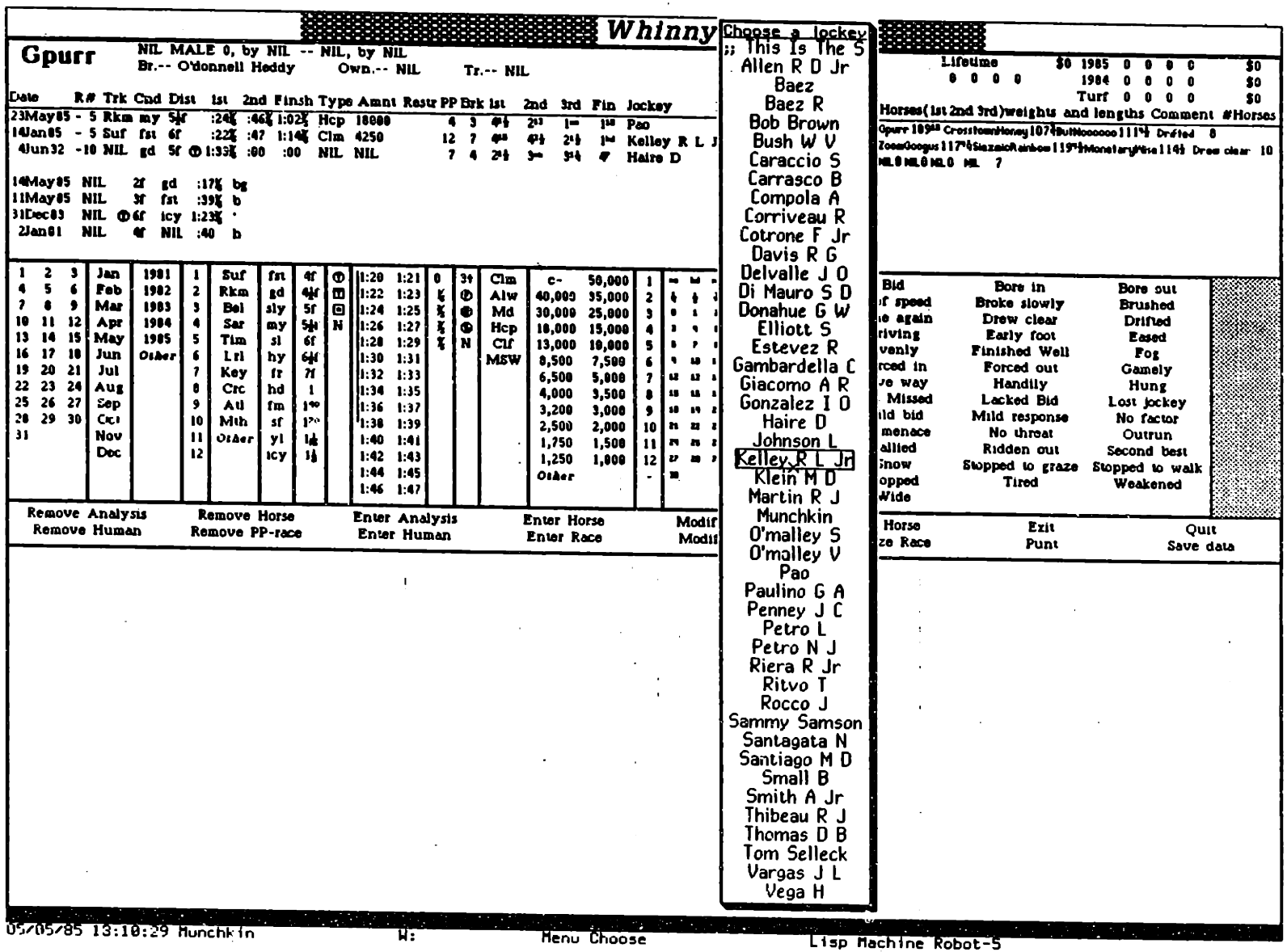


Figure 5: Example of a Temporary Menu (for jockeys)

like, please refer to Figure 5, for an example of the temporary menu for "Jockey" selection.

After the Track menu is the "Distance" menu, follow by the "Track Type, and "Track Condition" menus. After these were two menus, called "Call Times", and "Call Time Fractions". These menus were used to enter the times for the first call, the second call, and the final call. It was necessary to have the contents of the "Call Times" menu variable, as these times were related to both the field needed, and the distance of the race. When used to enter the first, second, and final calls, the "Call Times" menu would first display the list of the reasonable times for a first call, at the distance of the race being entered. After a first call time was selected from this menu, the "Call Times Fractions" menu would highlight, asking for the fraction which corresponded to the first call time. The "Call Times" menu would now change it's items to those times suitable for a second call, at the race's distance. This would continue, until all of the three calls, and their fractions had been entered.

The next menus were those for "Race Type" and "Race Class". Following these were the "Position Calls", and "Lengths Calls" menus, which contained items useful for entering the race position calls for the past race. Next was a "Weight" menu, followed by a "Comments" menu, and a "Number of Horses" menu. These were all of the permanent menus displayed in the *Whinny Frame.*

Entry of a horse and it's past performances

Data entry occurs in the following way. After a *Whinny Frame* is created, the command menu, is highlighted, i.e. the program is waiting for a command. When the command **Enter Horse** is clicked on by the mouse, the program enters a routine, prompting for, and accepting data for a horse entry. First the user must enter the name of the horse. The program then checks to see if it already has information about this horse. If so, it displays it, for the user, and asks the user whether he wants to modify the

horse, or just reuse the name and start from scratch. If no information on that horse is found, the WHINNY input creates a new horse, with that name. It then prompts the user for the horse's color, sex, and age, by creating temporary mousable item menus for each of these fields. Then the user must type in sire, dam, and grandsire. After this, the program prompts for owner, breeder, and trainer, in each case creating a temporary menu of all of the people of each occupation it already has in the data base. If the correct owner, breeder, or trainer is not listed on the menu, the user may simple move the mouse off of the menu, causing the menu to disappear, and then enter the name of the person through the keyboard. This name is then stored in the list of people of that particular occupation, and will then be offered on the menu, next time, as data for that particular field is needed.

Next, the WHINNY input prompts for the earnings, which are divided by lifetime, this year, last year, and turf. For each of these, the program individually prompts for number of starts, wins, places, shows, and amounts won, and the user must type in the data. *[6]*

Following this, the WHINNY input asks the user if he has any past performance data to enter for this horse. If "Yes", the program enters the routine for prompting and receiving past performance data. After each past race is completely entered, the program once again queries the user if there is another past race to enter. When the answer to this is finally "No", the input program then asks if the user has any recent workouts. These are prompted for and received until the user finally answers "No", at which point, a complete horse entry should exist for that horse.

Built into the WHINNY input system are a set of defaults for every field of a horse or a past performances. When the program prompts for an input, it also includes the default in it's prompt. If the user wishes to select the default, he merely hits a carriage return, which causes the default to remain the value of that slot. After data has been entered for a slot, the default for that slot of the current horse, or past

performance race changes to the current value of the slot.

While data items are being entered for a horse and/or it's past performances and workouts, the display pane is constantly updating itself to display the current existing information on the horse. In that way, the user can keep track of what he has entered, and also detect any error in the input. Should he find any, he can immediately interrupt the input cycle, and fix them, thanks to the following feature.

When the display pane displays information about a horse, or a past race, or workout. it creates a mouse-sensitive item for each piece of data in every field. The item contains, not only the printed representation of the contents of the field, but also the name of the data field, and the horse the data belongs to (past race and workout mouse items also contain information as to which past race, or workout they belong to). At any point in any routine for prompting and receiving data, the user may click on an item in the display pane instead. This will cause the program to drop everything, and re-prompt for the data for that field highlighting the proper menus or creating the correct temporary menus needed for entering data in this slot. After the new piece of data has been secured, the program returns to it's original location in the data entry routine, and re-prompts for the data field which it was on before the interruption again.

Time Saving Features in the WHINNY input system

The WHINNY input program employs various time saving schemes, designed to prevent the user from having to enter data which the program might be able to figure out. One of these schemes occurs in entering data for the horses which came in first, second, or third in a past race. If the final position call is 1, 2, or 3, that indicates that the horse whose past race we're entering was was either the first, second, or third horse. In this case, the program already has the horse name, and the weight carried. It may also have the lengths til the next horse, if the current horse was the winner! All of this is checked for, and valuable keystrokes saved!

As mentioned previously, the WHINNY program also maintains information about the people it encounters in it's input routines. If a name was entered for an "Owner" or "Jockey" slot, the name is saved in the data base, along with the information that this person is an owner, or a jockey. When it comes time to create a temporary menu for user selection of a person-filled field, the program rounds up all of the people with the necessary attribute, such as those people who are jockeys, or those people who are owners, etc. This list of people with the necessary attribute become the choices on the temporary mousable item menu. Additional people can also be entered into or removed from the database, along with their occupations, for this purpose, using the commands **Enter human**, or **Remove human**, available from the command menu in the *Whinny Frame.*

One other time-saving technique for data entry is also used. When the user enters the date, track, and race number for a past race, these specify a unique past race, in which anywhere from six to twelve horses might have participated. Many of the data fields in this past race will be the same for all of the horses in that race: track type, track condition, distance, first call, second call, final call (these are the times for the leader of the pack!), race type, race class, track variance, number of horses, and the first, second, and third horses, along with their respective weights and lengths won by. All of this data can be copied from one horse's past race to another's, and then only a smaller subset of the total fields prompted for. The scheme works best, of course, when there is a lot of data already in the data base, and the chances that a horse with a past race similar to the past race you're about to enter become very high. Thus, after a past race is uniquely specified, the WHINNY input searches the data base for a past race describing the same race, and copies all of the applicable data. Then it enters a modified input routine, prompting only for the remaining subset of data, unique to this horse in that race.

There are three other WHINNY commands, in addition to those previously mentioned, which are useful in manipulating horses. **Remove Horse** presents the user with a temporary menu of horse names, which when clicked on, will remove the horse from the database. **Modify Horse** also presents the user with a temporary menu of horse names. The horse corresponding to the selected name is displayed on the display pane, at which point it becomes easy to modify any data field, merely by clicking on them with the mouse. When this is done, the program behaves as it would if you were attempting to correct an error while in data entry mode. The correct menus either highlight, or appear, and the selected data is entered into the appropriate field. **Feed Horse** is a command which, in a manner identical to **Modify Horse**, pops up a list of all known horses, and displays the one selected by the user. After this, however, the program enters it's routine for prompting for more past races, and workouts. This is useful for updating data on a previously entered horse, which has run more races since it's initial entry into the database.

Entering Races and Analyses

A race to be analyzed is entered in a similar fashion to that of entering a horse. When the user selects the WHINNY command, **Enter Race** a routine is entered which prompts the user for date, track, track type, distance, race type, race class, and number of horses, highlighting permanent menus or popping up temporary ones wherever possible. Then the race input routine prompts for each horse entered in the race, by popping up the menu of all known horses, and then prompts for the corresponding jockey, weight, estimated odds, and equipment. An expected track condition is also prompted for, which defaults to "fast" if not entered. The data for a race is also mouse-sensitive, so that it, too, can be easily corrected if a mistake is made while entering a race. For an example of a race displayed in the *Whinny Frame*, please refer to Figure 6.

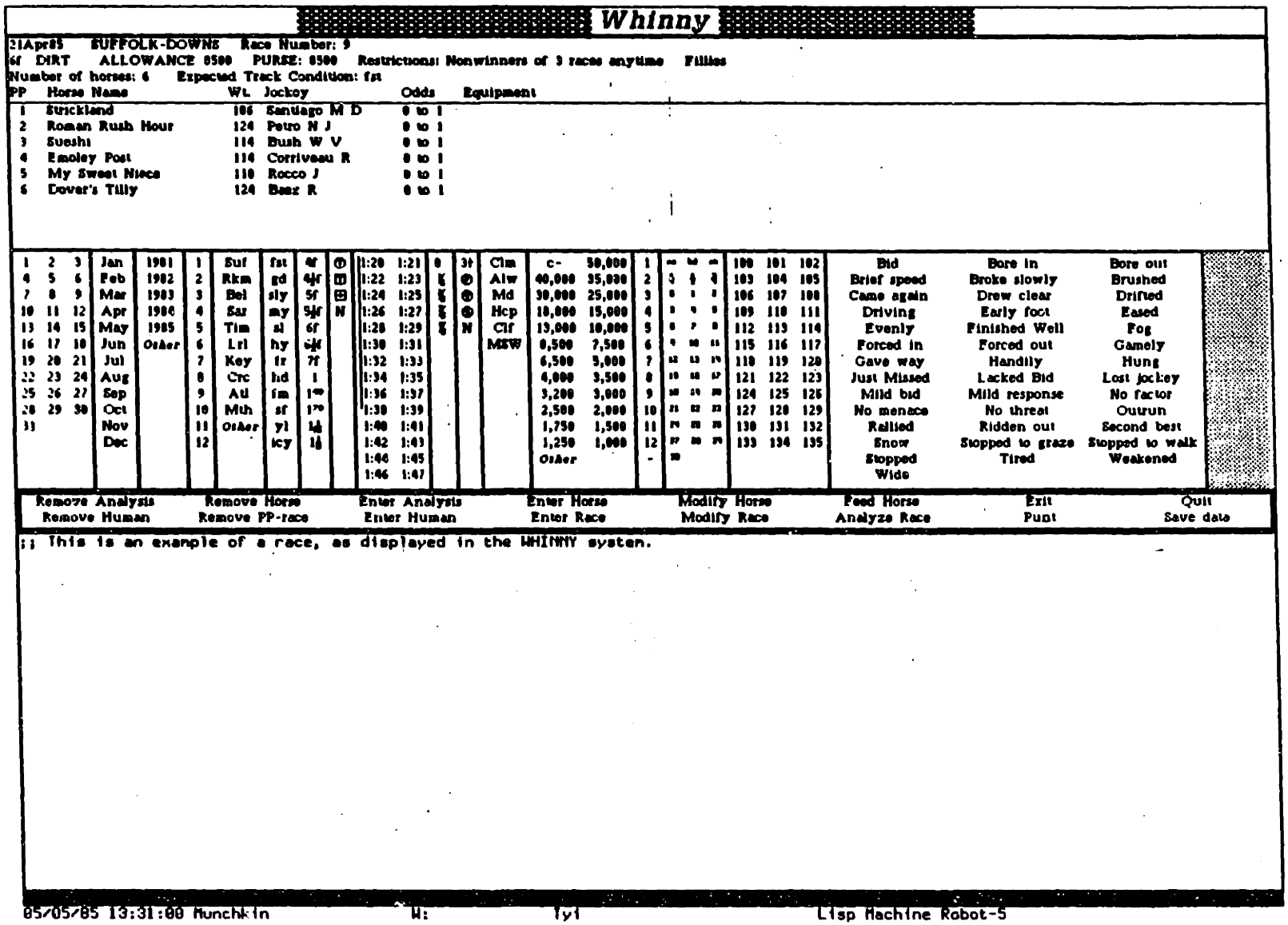


Figure 6: Example of a Race displayed in the *Whinny frame*

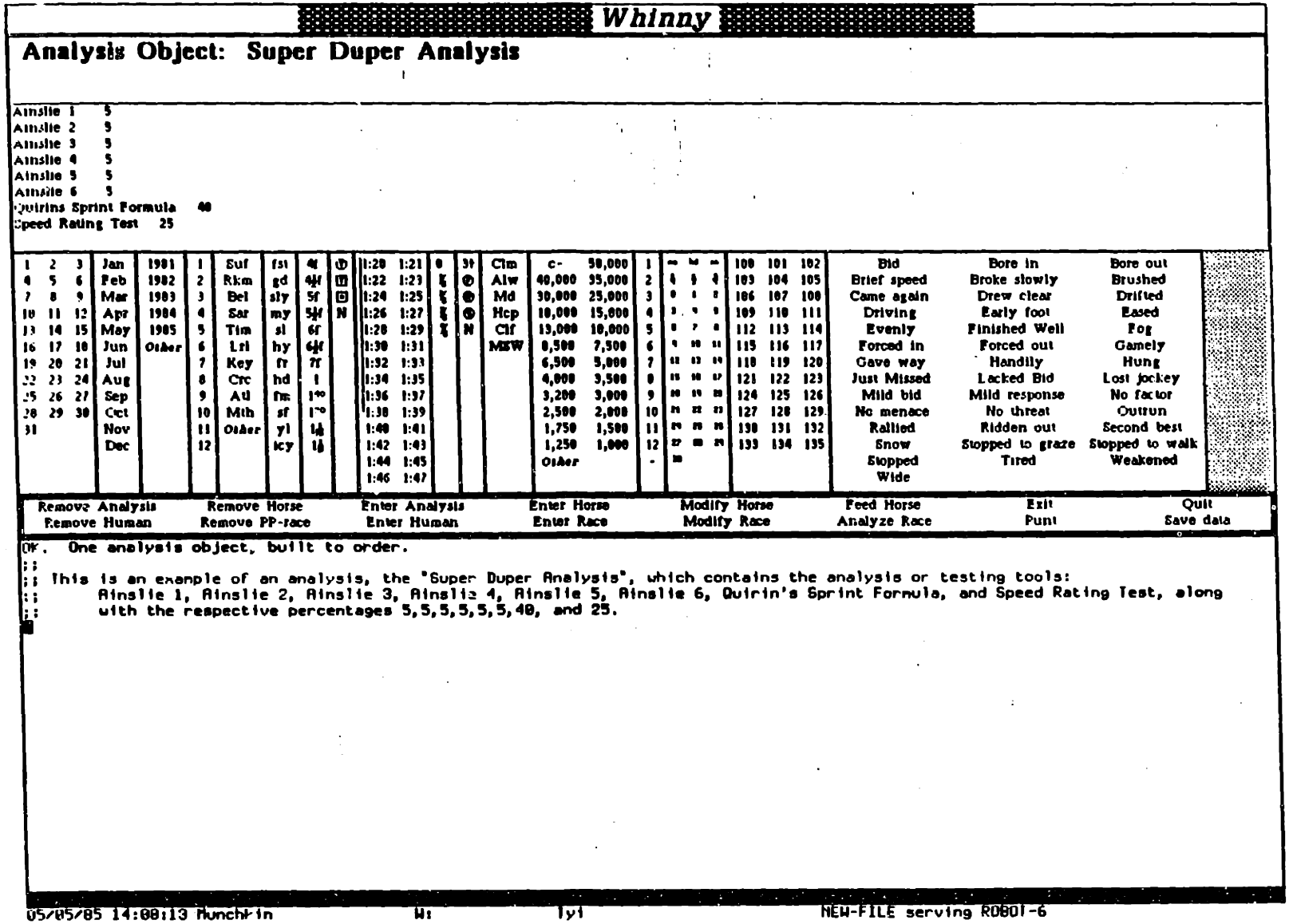


Figure 7: Example of an Analysis in the *Whinny Frame*

Entering analysis is also easy to do, in the WHINNY system. The WHINNY program keeps track of every rating scheme for comparing horses (often also called a "testing tool") ever programmed for use on a race, and every analysis created from rating schemes. When the command **Enter Analysis** is chosen, the program first asks the user for a name to describe the analysis about to be created. This can be any string with some meaning for the user, so that he can identify it, later. Then the analysis input routine pops up a temporary menu containing all known testing tools and analyses (those analyses already constructed). *[7]*

The user may then select those testing tools, and analysis objects he wishes to use in the new analysis which he is creating. After these are chosen, the WHINNY program asks the user for the percentages which he wants assigned to each testing tool, or analysis. These are the weights which will multiply the results returned by each testing tool, or analysis. In this manner, the user can experiment with which factors are more crucial to correctly handicapping a race. After all of the percentages are assigned, WHINNY saves the analysis in the database, where it can be later called on to handicap a race. For an example of the *Whinny Frame* displaying an analysis, please refer to Figure 7.

The actual workings of a WHINNY analysis will be describe in more detail in Chapter 5. At this point, however, I would just briefly describe the linear model which is currently being used in the WHINNY system. In this model, an analysis is comprised of a set of rating schemes or other analyses. Both rating schemes and analyses return a set of numbers, one for each horse, which indicate how well the horse scored on the test. In my linear model, this numbers are just weighted by some factor, and added together. The horse with the highest (or in some analyses, the lowest) *[8]* total score is the "best" horse, according to the analysis. As a quick example, suppose that in a analysis we considered "speed ratings" to count for 70% of the horse's success, and "jockey quality" to count for the other 30% percent. This analysis would consist of two testing tools (or

maybe two analyses, depending how we solved for speed ratings or jockey quality), with the percentage points of 70 and 30, for the speed rating and the jockey quality, respectively. The results returned from the two tests would be multiplied by the percentage points and added together. The horse with the highest total points would be the "best" horse. While there are other possible models which could be used to describe how the factors fit together, the linear model, as suggested Dr. Quirin [Qurn79] is a "first cut' at the problem, and the model which I chose to explore. But more on this later in Chapters 5 and 6.

The command **Analyze Race** is the WHINNY command used to invoke a particular analysis on a race. First, the user is asked to select the race he wishes handicapped from a temporary menu of all of the races WHINNY knows about. After a race is chosen, WHINNY displays the race contents, and checks with the user to make sure that this is indeed the race the user wanted. Then WHINNY pops up a menu of the names of all known analyses, and asks the user to select one. After one analysis is chosen, WHINNY prompts the user as to whether it should analyze the race. If the respond is "Yes", then the WHINNY system applies the chosen analysis to the selected race, handicaps the race, and prints the results on the display pane.

In this chapter I have described how the WHINNY system handles data entry, with the aid of mousable item menus. I have also mentioned some of the time-saving features built into the WHINNY input system to save time by eliminating the entry of redundant data. Finally I briefly mentioned the entry of races and analyses, with a small discussion about the model I will be using for analyses.

4. Rating Schemes for Handicapping Thoroughbred Races

Once I had programmed the framework for handling handicapping analysis, I needed some ways to compare the actual thoroughbred horses which would be competing against each other in a race. The regression formulae computed by Dr. Quirin [Qurn79], as mentioned in Chapter 2, would provide a solid basis, as they represented the result of a standard, mathematical analysis on a very large sample of horseraces. However, in the art of thoroughbred handicapping, more than just a glorified average is needed. The subjects of the study are horses; animals which suffer setbacks after strenuous efforts, and work themselves into condition through the very races they run in. Much of their readiness is also based on the competence of their trainers. These same trainers also have unique styles of working horses into top performance, while juggling important factors such as money management, and the success of the stable, in addition. All of these different, important factors become averaged out, however, in a survey that only looks to see what all the winning horses have in common!

Thus it becomes necessary to introduce more than just an regression formula to the analysis. The "rules of thumb" provided by Tom Ainslie in *Ainslie's Complete Guide to Thoroughbred Racing* just happen to cover some of the missing aspects of the problem. [Ains68] Each of Ainslie's "rules" presents a very specified situation in which one horse clearly has an edge over the competition. While the lucky trackgoer may find at most one or two horses a month which actually satisfy a particular rule, the use of all seventy-seven of the handicapping rules provides the handicapper with a sufficiently large number of plays per day. Some of the angles these rules of thumb take into account very from consistency, and speed ratings, to owner finances, and trainer styles. These rules provide the perfect complement to Dr. Quirin's regression formulae, as they pick out the type of the specific qualifications which the regression formulae have have averaged out.

Here, in this chapter, I shall present the rating schemes offered presented by Dr. Quirin, and Tom Ainslie for handicapping thoroughbreds. Then, in chapter five, I will describe how I put together four separate analysis, using both Quirin's and Ainslie's systems, first separately, then together.

Dr. Quirin proposes two separate handicapping systems: the sprint formula, for races under one mile in length, and the route formula, for races one mile and longer. In each formula, each horse begins with 2000 points. Then points are added for bad qualities, and taken away for good one. The horse with the lowest number of points is the best one, according to that analysis. [Qurn79]

Dr. Quirin's Sprint Formula

The Sprint Formula consists of seven parts. For each, I will give a short description of how it works, and how many points it is weighted by. I will also include the range of possible values, before the weighting factors is applied. I would have included the standard deviation, also, except that it varies greatly among different classes of horses, and different styles of races (sprint vs route), and as just an average number across all classes and distances, it would have very little meaning.

1. Days Since Last Race, Weight: +2, Range: 1 day to many years

This is a negative factor, as a horse loses condition and form after being away from the races. It may also indicate trouble with the horse, such as an illness or injury, which kept the horse from the track. For every day the horse has been away from the track, add two points. *[9]*

2. Number of Good Races in Last Ten Starts, Weight: -40, Range: 0 to 10

This is a positive sign. A good race is defined by Quirin as one in which the horse came in either first, second, or third, or finished within two lengths of the leader in sprint race, or three lengths in a route. This indicates that the horse has some sign of life in it, and that it has been at least within striking distance of the winner of that race.

3. Number of Failures, Weight: +37, Range: 0 to 8 or higher *[10]*

A failure exists when a horse has not had a good race against either the same company and distance which it will be facing today, or against the same company and distance as it's last good race. In either case, the horse has failed from what should have been expected of it. Thus these points are added, to increase the horse's total number of points.

4. Speed Point Percentage, Weight: -14, Range: 0 to 8

It is a generally accepted truth that the horse that gets out ahead of the pack early in the race is less likely to run into trouble, such as being boxed or brushed, in the race. Thus, the racing "luck" of a speedster from the gate is increased. Dr. Quirin presents a complex set of requirements as to how many "speed points" a horse should receive per race, and under what conditions. The general idea behind it all is to give points to the horse which most often leads, or is near the lead of the pack at the first call in the race. These positive-factor points are subtracted, to lower the horse's total points. *[11]*

5. Rank: Average Earnings per Start, Weight: +93, Range: 1 to 5

Horses are ranked by how much money they have earned per start. The fact that this is done per start is to even out the differences between good, young three-year-olds who may have less money, total, due to their short careers, than more mediocre old horses

with long, unsuccessful careers. Horses are "ranked" in by giving one point to the horse which is first in his field, two points to the second horse, three points to the third horse, four points if otherwise in the front half of the field, and five points to those horses in the back of the field.

6. Rank: Average Speed Rating Last Two Good Races, Weight: +116, Range: 1 to 5

A horse's speed rating is a number which represents how fast the horse has run in a race. It is usually thought of as the sum of two other numbers: the first being the actual speed rating, and the second, the track variance. (These terms are explained in the Handicapping Primer at the end of Chapter 1.) The actual speed rating is the number which results from: 1) Calculating the difference in fifths of a second of the horse's finish time from the time of the track record for that distance, and 2) Subtracting one point from one hundred points for every fifth of a second of the difference. This results in a number, less than one hundred, which is smaller for slower horses, and larger for faster horses. This, however, is not enough, due to variances in the track surface. Most horses, not matter how good, would have trouble coming anywhere near the track record on a muddy track. *[12]* Due to varying track surfaces, a number called the "track variance" is calculated, which roughly is the average of the differences from the track records of all the winners on any given day of racing. Thus, on a muddy day, all of the final times would be slower than the respective track records, and the resulting track variance would be a large number (say, around thirty points, or so). On a faster day, the track variance would be lower, (say, around fifteen points).

To present an accurate total speed rating, the two numbers for the actual speed rating, and the track variance, are added together. Even this number need modification, however, due to the following reason. At many minor tracks, such as Suffolk Downs, the track records are set by horses of much higher quality (usually visiting), than the normal animals which frequent this track. Thus, on a normal day, all

of the Suffolk horses have actual speed ratings must lower than the track records. This results in a large track variance even on fast days, with an even larger variance on slow days. The final result, is that, in comparing Suffolk horses with horses from other, better tracks, the Suffolk horses look like they have been running slow times on very bad tracks, when actually, they're just slow. This would cause handicappers who rely on the total speed ratings to overestimate the Suffolk horses, when comparing them to other horses shipping into Suffolk Downs.

Dr. Quirin realized this phenomena, and devised a way to compensate for it. By keeping track of the final times for horses from different tracks, at different distances, he was able to create a chart of numbers, which, when added to the total speed rating, created a set of figures which were consistent across different tracks. This "Modified Quirin Speed Rating" is what I have used to calculate the speed ratings used in this test.

This this test, the modified Quirin speed ratings are calculated for each horse's last two good races. These are averaged together to get a more accurate number. Then the horses are ranked, in the same manner as for the "Average Earnings per Start'. Horses receive one point, if first in the field, two points if second, three points if third, four points if otherwise in the front half of the field, and five points if in the back half of the field. These points, multiplied by 116 (the weighting factor) are added to the horse's total points.

Jockey Rating, Weight: +250, Range: 1 to 3

Quirin devised the following rating system for jockeys. One point is given for a top five jockey. This is a jockey who is one of the five best jockeys at the track, as noted by his win-percentage. Trainers often spend the extra dollars needed to get one of the top five, when they are ready to spring the horse for one of his best races. Hence, a change of jockey, from a mediocre one, to one of the top five, is something handicappers watch for. Two points are given to any jockey who has been previously successful on the

horse, i.e. has had a "good" race on the horse. At least this proves that the horse will perform in a good manner for this jockey, and that the pair has been successful in the past. Not a bad bet. Three points are given to any jockey who does not fit in the first two categories. These points are multiplied by the weighting factor of +250, and added to the horse's total. *[13]*

Those are the tests used in Dr. Quirin's Sprint Formula, and used by my computer simulation of his analysis. As stated before, each horse begins with 2000 points, and points are added or subtracted as indicated by the results of each test. The horse with the lowest final number of points, is theoretically the best horse.

Dr. Quirin's Route Formula

The Route Formula consists of five parts. For each, I will once again present a short description of how it works, how many points it is weighted by, and the range of values returned by the tests before the weighting factor is applied.

1. Post Position, Weight: +74, Range: 1 to 13

In a route race the starting gate is much closer to the first turn than in a sprint race. Thus, for a race of this distance, the horse which has the inside rail for the turn has an advantage over those horses which have to traverse the turn three or five horse-widths away from the rail. Unless there is a noted superstar fast-breaker in the route (in which case, after such an enormous burst of energy, it is questionable if the horse could last the long distance), the horse with the inside post positions will have the short distance around the turn. *[14]* Post position one is on the inside rail, with the higher post positions working out toward the outside rail.

2. Number of Wins in Last Ten Starts, Weight: -11, Range: 1 to 10

One of the best indicators of how well a horse may do, is the results of how well he has done in the past. In this simple test, the number representing how many times the horse has won in the last ten starts is multiplied by the weighting factor of 11, and subtracted from the horse's total points.

3. Jockey Rating, Weight: +228, Range: 1 to 3

This is calculated in the same way as for Dr. Quirin's Sprint Formula. One point is given for a top five jockey. This is a jockey who is one of the five best jockeys at the track, as noted by his win-percentage. Two points are given to any jockey who has been previously successful on the horse, i.e. has had a "good" race on the horse. Three points are given to any jockey who does not fit in the first two categories. These points are multiplied by the weighting factor of +228, and added to the horse's total.

4. Rank: Average Earnings per Start, Weight: +115, Range: 1 to 5

This test is also identical to that used in Dr. Quirin's Sprint Formula. Horses are ranked by how much money they have earned per start. Horses are given one point for being first in his field(i.e. having the most money per start, compared to the other horses in the race), two points if second, three points if third, four points if otherwise in the front half of the field, and five points if in the back of the field.

5. Rank: Average Speed Rating Last Two Good Races, Weight: +110, Range: 1 to 5

This test is once again identical to one used in Dr. Quirin's Sprint Formula. A modified speed rating is calculated for each horse, using. Quirin's method of modification, using the last two good races of the horse's past performances. These two speed ratings are averaged together, to get a truer idea of the horse's ability. The modified speed ratings of all the horses are then ranked, as explained before, and the resulting numbers

weighted by +110 before being added to the horse's total number of points.

Those are the five tests specified by Dr. Quirin's Route Formula, and used by my computer simulation of his analysis. As stated before, each horse begins with 2000 points, and points are added or subtracted as indicated by the results of each test. The horse with the lowest final number of points, is, once again, theoretically the best horse.

While I realize why Quirin used "ranking" on the results of the average speed ratings and the average earnings per start, I also feel that is results in the loss of some very significant data. The ranking is used to keep the various factors in the test in "proportion" with each other. Suppose, for example, that $3000 class horses had an mean "average earning per start" of only $100, while higher class horses, say $10,000 had a mean "average earning per start" of $500. If just this numbers were weighted, then this factor would count more for higher class horses, and less for cheaper horses. So ranking must be done, to account for these differences. Or must it? Suppose that the average earning per start was "normalized" by the class of the horses under consideration. By doing this, we could use the raw normalized numbers in the linear formula, instead of "ranking" the horses and assigning them numbers 1 through 5 for how they "compared" to the other horses. When using ranking, some very important differences are smoothed over. Suppose that the horse with the highest earnings per start is really much higher than the rest of the field, and the rest of the field is roughly equal (plus or minus a few dollars). This important fact is wiped out by a ranking that just returns numbers 1, 2, 3, 4, and 5. The same is also true for speed ratings. All of the horses may be of a similar speed, or one or more may be superior, but the ranking just returns numbers 1 through 5. I believe that this is the weakest step taken in Dr. Quirin's formulae, and an unnecessary one. If the results returned by the "ranked" tests were instead "normalized" to some quality of the entire field, such as the class of the race, they could still be used in the linear formula without disrupting the weightings, but with much greater accuracy.

Now that I have explained the tests used by Dr. Quirin to generate a handicapping system which is based on the average results on many races, I will present some of the rating schemes, or "rules of thumb" proposed by Tom Ainslie in *Ainslie's Complete Guide to Thoroughbred Racing.* These schemes describe racing situations which occur infrequently, but which also provide excellent betting opportunities to those handicappers (or computer programs) who look for them. [Ains68]

While Tom Ainslie presents seventy-seven different angles for play, I shall only include a few examples, so that the reader may understand the flavor of these rating schemes. A good computer analysis, however should include all seventy-seven of them.

Ainslie's System 1

Play a horse entered for a claiming price at least $2,000 below a price tag it has carried in the past, provided it is the only such horse in the race, and is regarded highly enough to go postward at odds of 15 to 1, or less.

Explanation: Although the decreasing market value of the horse indicates the possibility of problems, the horse is still facing competition of cheaper quality than that which is has recently raced against, and may be able to handle this level of racing.

Ainslie's System 2

Play any horse running in a claiming race for the first time, provided it has been racing and working out regularly, was beaten by not more than five lengths in its last start, and either led or was within two lengths of the leader at some stage of the race.

Explanation: The claiming class represents a lower quality than the allowance classes the horse has been recently running against. This compiled with the fact that the horse had some life in it's last race, indicates good possibilities for this race.

Ainslie's System 3

Play a horse that led or ran within a length of the leader at one or more of the early calls of its last race and drops down today to a class at which it has won in the past.

Explanation: Once again, a sign of life in the last race, coupled with lower quality competition in this race.

Ainslie's System 4

Bet any horse that was claimed in one of its last three races and is entered today at a price lower than the new owner paid for it.

Explanation: "Horsemen sometimes make dreadful claims, but not very often. If the horse has been running and working regularly and the owner now risks losing it for less than he paid, the player knows that the barn wants today's purse. Badly." [Ains68]

Ainslie's System 5

Play a horse stepping down in class after a race in which it ran out of the money, provided that (1) it had stepped up in class for that race, and (2) it was never worse than third at any call in its next-to-last race.

Explanation: If the horse had a good enough next-to-last race that the owner attempted to step him up in class, then he should be easily able to handle the original class once again.

Analyses and Tests	Range	Mean	S.D.	Percentages and Weights
Quirin's Sprint Formula	unbound	11.16	3.62	50%
Quirin's Route Formula	unbound	14.11	2.91	50%
Ainslie's Rules of Thumb	——	——	——	50%
Days Since Last Race	1 – ?	34.6	49.5	+2 S
Number Good Races in last 10	0 – 10	3.36	1.34	-40 S
Number Failures	0 – 8	.71	1.28	+37 S
Sprint Speed Point Percentage	0 – 8	4.36	3.33	-14 S
Rank: Average Earnings per Start	1 – 5	3.45	1.43	+93 S +115 R
Rank: Average Speed Rating	1 – 5	3.00	1.30	+116 S +110 R
Jockey Rating	1 – 3	1.95	.49	+250 S +228 R
Post Position	1 – 13	3.5	1.71	+74 R
Number of Wins in last 10	1 – 10	1.5	.77	-11 R

S: Weight in Quirin's Sprint Formula
R: Weight in Quirin's Route Formula

Table I: Tests and Analyses of the WHINNY system

These are examples of some of the betting rules proposed by Tom Ainslie, and some of the algorithm which my computer analysis uses to test for the "best" horse. [Ains68] As stated before, these specific cases serve to round out the more general analysis method proposed by Dr. Quirin. Both methods of analysis used together should cause an increase in profits over either of the systems used separately. For a total listing of all of the testing tools and analyses currently in the WHINNY analysis system, along with their ranges of values, means, and standard deviations, please refer to Table I.

There is one more item which falls under the category "rating schemes", though not directly. Once we have completed an analysis, how should we decide to bet? What are our rating schemes for betting strategies?

The analyses in the WHINNY system return results in the form of numbers, one for each horse, which try to predict the overall quality of each horse. Given these numbers, how can we decide how much money, and in what type of bet, to play? Probably the best way to decide this it to handicap a large number of races, along with their results and payoffs, and then determine how much money would be made if different bets were made. Some reasonable bets could be 1) Bet win on the "best" horse, 2) Bet place on the "best" horse (cautious betting), 3) Bet show on the "best" horse, 4) Bet an exacta on the "best" and "second-best" horses, 5) Bet combinations of exacta with the "best" horse first, and the next 3 best horses as second (3 different bets), 6) Bet a trifecta on the three top horses in the order the WHINNY system predicted, 7) Bet a trifecta on the three best horses in any order (6 different bets, but they cover all combinations of these horses), etc. All of these betting strategies are sound ones, yet some may be better than others, depending on the situation. For example, if the WHINNY system shows that one horse is clearly superior to the rest of the field, then betting that horse for win might be the best thing to do. If the WHINNY system returns three outstanding horses, then maybe a trifecta of these three horses in any

order would be the smartest bet. (Sure, you could bet "win" on all three horses, but then you're guaranteed at least two loses, and if the favorite wins, it may not pay back enough to cover the two losing bets.)

The WHINNY system could be used to suggest betting strategies, basing on the point values returned by the analysis used to handicap the race. I have not conducted any testing of betting strategies at this date, however I feel confident that correlations do exist between the point values of the handicapping analysis, and the correct betting strategies which could be used to maximize profits.

This chapter described the rating schemes which I will be using in the WHINNY analysis system. These schemes include Quirin's Sprint and Route Formulae, and Tom Ainslie's "rules of thumb". In addition, I briefly mentioned the possibility of having betting schemes, and how WHINNY could be used to suggest appropriate betting schemes for different situations.

5. WHINNY System Analysis

Back in Chapter 3, I briefly described how the WHINNY system structured an analysis, and how one was entered through the WHINNY input system. In this chapter, I will describe how the WHINNY System actually carries out the analysis of a race. This will involve first describing how an analysis is put together, in more detail, then how one is invoked, and with what arguments, and finally, how the results may be interpreted. After this will come a short presentation of the four different analysis methods which currently are implemented in the WHINNY system, and how they work.

An analysis is created out of either primitive tests or other analyses. During the creation of the analysis, each of it's sub-analyses and tests are assigned a weighting percentage. These weights will be used to modify the results returned by the separate analyses or tests, before they are added together. Thus, a WHINNY analysis is merely a list of either primitive tests (programmed in the system), or other WHINNY analyses, and their respective assigned percentages. For an example of the hierarchical nature of analyses, see Figure 8.

The system is generalized by the fact that both WHINNY analyses and the primitive tests are invoked in the same way, and return answers in the same format. The implementation of both the analyses and tests are as "flavors", which is a type of object that may be programmed in ZetaLisp, on a *Symbolics 3600* Lisp Machine. One important feature of flavors is their ability to be invoked by sending them a message. Different flavors may do different things on the receipt of the same message, depending on what they have been programmed to do when they receive that message. For example, the menus which I have been using to select my data from are "instances" of a flavor of "menu". An instance is one realization of a flavor (in the same way that, for example, Sam the collie, is one instance of "dog".) When the WHINNY system sends a menu the message "Highlight-yourself", the menu becomes highlighted, thus attracting

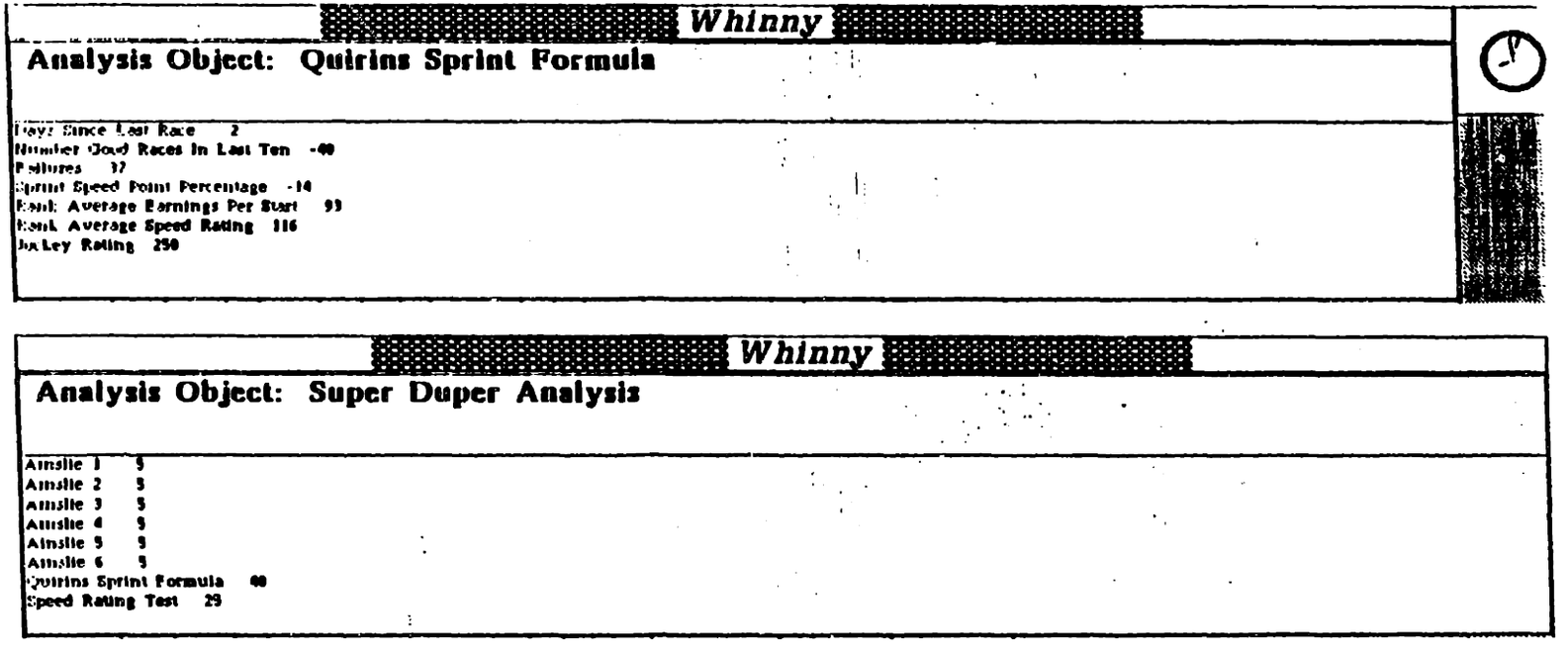


Figure 8: Example of an Analysis

Note: the hierarchial nature of "Super Duper Analysis" which contains "Quirins Sprint Formula" as one of its components

the user's attention.

In the WHINNY system, both analyses and primitive tests are invoked by sending them a message "Handicap-Yourself", along with the race to be handicapped. Remember that a race includes all of the necessary information, such as the data, track, distance of the race, and all of the horses, and their respective jockeys and weights. One more feature is important about a race. Upon construction, the horse's in the races are sorted by post-position number, and ordered from 1 to n, where n is the last post-position in the race. This order is preserved by the tests and analyses receiving the race, and it is the order in which the results are returned for each horse. Thus, when the points are combined by the parent analysis for each horse, there is no confusion or mixup about which points belong to which horse.

When an analysis receives the message "Handicap-Yourself" and a race, it merely sends this message and the race to each of it's sub-analyses and sub-tests. Each receiving analysis, in turn, passes the message and the race to each of it's sub-units, until at last, at the bottom, only primitive tests are receiving the message. Unlike their callers, the analysis objects, these primitive tests actually do something useful with the race. Let us consider the simplest case, which is an analysis object composed only of primitive tests, and start our discussion there. This analysis receives the message "Handicap-Yourself" and the race, and dutifully passes the message and the race to each of its tests.

When a test receives the message "Handicap-Yourself" and a race, it executes some piece of code which tests one particular factor which would influence the race. When, for example the Quirin test "Rank: Average Speed Rating in Last Two Good Races" receives the message, it determines the average speed ratings in the last two good races for each horse in the race, from their past performances (which are included in the race argument). Then this test ranks the horses by their average speed rates, normalizes the results, and returns a list of numbers, representing the normalized

ranking points given to each horse. The numbers return in the same order as the horses appeared in the race, which, as mentioned previously, is the standard order that all of the tests and analyses will return their results in.

The normalization is required because each test may not necessarily return values in the same number range as every other. Only a test knows how to normalize it's own results, because only the test itself knows what possible values can be returned. Analysis can also be informed how to normalize themselves, although this is a little trickier. First, let me demonstrate why normalization is necessary. Say, for example, we wanted to weight two tests equally, and gave them each a weighting factor of 50%. Suppose the first test returned answers in the range of 5 to 10, while the second test returned answers between 100 and 200. If we merely multiplied the answer returned from each test by 50% and added them, the second test would still carry more weight then the first, due to it's larger numbers. Thus we normalize each test to 100 points, such that the first test would return evenly between 0 and 100, and so would the second test. Note that no normalization is required with Dr. Quirin's tests, because these differences in results have already been accounted for. When combining Dr. Quirin's test, or analysis with other tests and analyses, normalization must be done.

Normalization of the results returned by a test is a standard part of the test's calculation. When a test is programmed, two variables inside the body of the test, call "high" and "low" are assigned values by the programmer. These are, respectively, the highest and lowest values which the test can ever return. The formula for normalization of each value returned by the test is the following:

((Value - Low) * 100) / (High - Low)

This is the current value minus "low", then the quantity multiplied by 100. This is then divided by the quantity resulting from "high" minus "low". The result of this division is the normalized value which is returned to the parent analysis object.

Analyses normalize their results in a similar manner. First, however, they need to request from each of their sub-test and analyses their respective highs and lows. (This can be easily accomplished by sending each of the tests and analyses the message "Highs-and-Lows". Flavors are really useful for message passing.) When these are received, the parent analysis weights each high and low by the weighting for the respective test from which it came. Then all of the highs are added together to form a new "high" for the parent analysis. Similarly, the lows are weighted and combined to form the "low" for the parent analysis. These two values are now the high and low which the parent analysis uses to normalize its results. The normalization is carried out in the same manner as before. Note that this process is consistent no matter how many layers exist in any analysis.

Thus, each test returns a normalized list of numbers, one number for each horse in the race. These numbers are all multiplied by the respective weighting factor assigned to the test when the result is received by the parent analysis object. The analysis object then adds up all of the weighted results of each test to come up with one final list of numbers, still one number per horse. These represent the final points given to each horse for the race. Dr. Quirin has chosen to assign percentages and point values, such that the horse with the lowest number of final points is the best horse. When using Tom Ainslie's tests, however, the horse with the most final points is the best horse. This is taken account for, and "normalized" when the two methods are used together.

There is really not much difference between this type of simple analysis, which we have just discussed, or a more complicated one, composed of other analyses, as well as primitive tests. Let us quickly consider an analysis composed of both other analyses and tests. When this composite analysis receives the message "Handicap-Yourself" and the race, it dutifully passes the message and the race to each of its components. If the component is a test, it handles the message and the race, by executing it's specific test

on every horse in the race, and then normalizing and returning the result. If the component is another analysis, however, it merely passes the message and the race to each of *its* components, however many they may be. When the results come back, and are weighted and added; then the final listing is normalized within that analysis object and returned to the parent analysis object. The parent analysis may also normalize it's results before returning them to the user, if this is requested. The user is asked during the creation of the analysis if he wishes the results to be normalized. While normalization is really not necessary for the top parent analysis, it keeps the results between 0 and 100, which may be aesthetically pleasing.

Thus the results returned to a parent analysis objects are of the same format, not matter whether computed by a primitive test or another analysis object. In this manner, WHINNY analyses can be created, using any combination of existing analysis objects, and primitive tests, in order to easily try out all interesting combinations of handicapping methods.

There are four major analyses implemented in the **WHINNY** system to date. These are: Dr. Quirin's Sprint Formula, Dr. Quirin's Route Formula, Ainslie's Rules, and a "Quirin and Ainslie Combination Method".

The tests used in both of Dr. Quirin's formulae, Sprint and Route, have been described in the previous chapter. The analysis objects, Dr. Quirin's Sprint Formula, and Dr. Quirin's Route formula, are just wrapper for the respective tests to be used in each analysis. The results returned by each test are weighted by the appropriate weighting factor. Then, all of the weighted results are added together. The horse with the smallest number is theoretically, the best horse.

The analysis "Ainslie's Rules" is just a grouping of all of the rules describing good betting situations in which one horse stands out from among the rest. Each of the rules is programmed in the **WHINNY** system as a primitive test. Each test returns one point for every horse satisfying it's conditions and zero for each horse which does not meet

the conditions. A test may "abstain", i.e. return all zeros, if none of the horses satisfy of the test's conditions. The results from these tests do not need to be normalized, as each test returns, for each horse, either one or zero. The results of all of the tests are summed together, and the horse with the most points, is theoretically the best horse.

The analysis "Quirin and Ainslie Combined Method", is just an analysis built out of both of Quirin's formulae, and Ainslie's Rules. If the race is a sprint, the analysis "Dr. Quirin's Route Formula" just abstains from analyzing the race; in the case of the race being a route, "Dr. Quirin's Sprint Formula" abstains from the analysis. Whichever analysis is appropriate is assigned 50% of the weight of the analysis, with "Ainslie's Rules" making up the remaining 50%. The results from either of Quirin's formulae are normalized to return values in which the best horse has the biggest number, with numbers being between zero and one hundred (this is accomplished by just flipping the sign on the weighting percentages for each Quirin test.) "Ainslie's Rules" is also normalized to return a final value between zero and one hundred. The results from Quirin's and Ainslie's methods are weighted by 50% and added together. The horse with the biggest final number is theoretically the best horse.

Once again, this combination is just a "first cut" at the issue, and probably not a very reasonable one. First of all, the weightings of 50% are just an educated guess, and may prove to be wrong under testing. But an even more important issue is at stake here. While the two analyses, Quirin's Formulae, and Ainslie's Rules cover different territory, it seems that a linear combination is not the way to combine this information for the best advantage. Ainslie's Rules are designed to pick out advantageous betting situations. If one of these is found, it would appear to be the logical next step to check and see if the horse selected by Ainslie's rules will have any serious competition. If the Ainslie test is one in which the selected horse is one dropping in class, we know we have a quality horse. If it has raced recently, then the next thing to check for would be speed ratings. Will this horse have any serious speed competition in the race? One of the

things less important would be "average earnings per start". If this horse is dropping in class, it probably means that it wasn't winning in the higher classes. It does not mean that a lower class horse with higher "average earnings per start" is a better horse. This test is not really valid under these conditions.

What I am suggesting is a type of "expert system" approach to the problem, where the tests used to handicap the race are chosen on the basis of the results of previous tests. In this case, we would use "Ainslie's Rules" to direct the "flow of traffic" of which tests would be applied. Other tests could also be used at the top level. For example, right now, we used the "distance" test to choose which Quirin test to use. If the race is a route, we apply Quirin's Route Formulae, while if it is a sprint, we use the sprint formula. Tests for "class" would also be good for top level tests, as top class horses hold their good condition much longer, and therefore do not have to have "recent action" (a race in the last 14 days) as a qualifier as a contender. These are just some suggestions as to a better model for the problem. As will be shown in Chapter 6, the results from testing just the linear combinations of factors are really not too bad. Especially as a first cut. But I feel confident that the linear model is just the "tip of the iceberg", so to speak, and that a more complex model would have a even better success factor.

One more feature about analyses remains yet to be discussed. One of the requirements for my system, as mentioned in Chapter 3, was that I needed to be able to easily construct and test new analyses and combinations of analyses. I hope that from my discussions on the entry of analyses and their actual operation that this feature has become apparent. New analyses can be constructed on the fly in a matter of seconds, using the WHINNY input system from previously defined tests and analyses. Defining new primitive tests takes a little longer, as this must be done as code in ZetaLisp. But once this tests are defined, they may be used in any combination whatsoever, to analyze a race.

There is another useful feature of this system which turns out to be very helpful, although it was not thought of, initially. It often is the case, when trying to analyze a race that the jockeys for one or more horses are not known yet. This is rather important, as both of Quirin's tests base a heavy percentage on "jockey quality". The WHINNY system can be easily used to solve this problem. Jockeys come in three different types: 1) top-5 jockeys, 2) those which have ridden the horse successfully in the past, and 3) those either untested on the horse or with bad rides on the horse. When the jockey is unknown before the race, the user can make some guesses as to who the jockey may be. In many cases it is the jockey who rode the horse in it's last race. Or else it is a better "top-5" jockey. The user can use the WHINNY command **Modify Race** to set the jockey to one of these, and then handicap the race. If the user wants to know what might happen if a different type of jockey, say a one of the top-5, rides the horse, he can use **Modify Race** once more to change the jockey to one of the top-5 jockeys at that track. Then the user can just handicap the race once more. Now the user is ready for either situation, and can take both analyses to the track the next day, and depending on what jockey rides, use one analysis or the other. This is another example of how the WHINNY system can make changes in data "on the fly", and re-handicap a race to see the effects of the changes.

In this chapter, we have seen what the structure of analysis objects is, and how they may be called on to handicap a race. All of the component analysis objects and tests of a parent analysis are given the race to be handicapped, and a notice to handicap it. Each test and component analysis returns a normalized list of numbers, one for each horse, which represent how well that horse scored on that test. The final result of the parent analysis object is the sum of the weighted results from each component analysis or test. I have also described the four analysis methods which are currently implemented in the **WHINNY** system, these being: Dr. Quirin's Sprint Formula, Dr. Quirin's Route Formula, Ainslie's Rules, and Quirin and Ainslie Combined Method.

That latter of these analysis actually fulfills our goal of creating an analysis which took into account both the average conditions present in winners, and those specialized situations in which a winning horse can be spotted in a "non-average" way. This linear model may not, however, be the best combination of these two tests, and perhaps something resembling an "expert system" might be much more successful. In either case, the WHINNY system can successfully make changes to both races and analyses quickly and easily, and then re-handicap a race to determine the effects of these changes.

6. Testing and Results

This chapter describes how the WHINNY system, both input and analysis, stood up under actually testing. The first aspect of WHINNY which I tested was the input system. This, of course, was vital, as I could not analyze a race, until I had actually managed to enter one. After I had entered some actual horses which were running or had run, in real races out at Suffolk Downs, I then handicapped the races, using my analysis methods, in many different combinations. The results, so far, indicate that the WHINNY system is performing well enough to show a profit in the long run. Along with the monetary results of the testing, I also will be discussing in detail some of the other models which I would like to explore with the WHINNY system sometime in the future.

Input Testing

Earlier, I had mentioned that one of the biggest problems in creating a computer handicapping program would be entering the massive amount of data necessary to the analysis process. With this factor in mind, I tried to create the best user-friendly, guess-ahead, mind-reading input system possible, so that data entry could be accomplished in a reasonable amount of time. Unfortunately, so far, the tests which I have done indicate that WHINNY input system is still not performing fast enough to make the system useful for entering small quantities of data in a reasonable amount of time. For large quantities, such as those which would exist if the input system were used to update on a daily basis all of the horses which race at Suffolk Downs, the system behaves much more quickly, and the entire update can be accomplished in a very small amount of time.

After practicing with the input system for a few tries, I was able to enter horse entries into my database at the rate of one every twenty minutes. This is the speed for one person entering data. If a two person team is used (one person reading the data

aloud while the other person does the entry), the time per horse gets down to ~15 minutes. (Not as fast as two people separately. At the one-person speed, which is what I will discuss from now on, it was taking me two hours to enter a six-horse race, and four hours to enter a twelve-horse race. Given that there are ten races on an average day of racing, and that a race has, on the average, eight horses, this indicates that it would take thirty hours to enter a day's worth of races. Still not fast enough.

This is not the entire truth, however. Because the input system employs checks to see if a past race has already been entered for another horse, and if so, it copies all possible data to the past race which is currently being entered, the time needed for data entry decreases rapidly as more and more data is entered. Considering that, on the average, eight horses run in a race, this means that seven out of eight past races being entered in the input system will be redundant. The data which may be copied from one past race, already entered for one horse, to another past race, for another horse, comprises over half of the total data in a past race. What this means is that, out of eight past races, only one would have to be fully entered, and the others only "half" entered.

This effects the time needed for data entry in a drastic way. In the previous paragraphs, horse entries took 20 minutes to enter. A past race takes ~2 minutes, and 15 seconds to enter from scratch. A past race which has been found in the data base only requires ~ 1 minute to enter the remaining (uncopied) data. Thus, we are saving 8 minutes, and 15 seconds on every horse entry we enter, and the total time required for a horse entry is down to only 11 minutes and 45 seconds. The time needed to enter a whole race is down to 1 hour, and 34 seconds. A whole day's worth of racing can be entered in just 15 hours and 40 minutes. (Plus the time needed to enter caffeine into the user who is entering all of this data).

Not bad, but it gets even better! Those were the figures for the times needed to enter horse entries from scratch, given that we have a full data base, (in which we can find seven out of every eight past races that we try to enter.) If this is the case then we

are no longer entering horse entries from scratch! If we are maintaining our data base up to date and including all of the horses which usually run at the track under study, then the user should only have to enter one past race for every horse in a race. These would be the past races for the last race which the horse ran in. All other past races would have been entered already into the data base. Thus, for a race, the user would only have to enter one past race for every horse, out of which seven of eight past races would be found by our checking system, and only require "half" entry. At this rate, it would only take 9 minutes and 15 seconds to update the existing information in the data base to the point where a race could be handicapped. Thus, one could successfully update the data base for a day's worth of races (ten) in only about an hour and a half!

The first goal of the WHINNY system had been realized. At the cost of only an hour and a half per day, a devoted handicapper could easily keep his database up to date. (Actually, as most tracks only race at maximum five days a week, database upkeep would really take even less time.) Unfortunately, those of us who are not quite so devoted, are, as some would say, still "up a creek." Without a complete, updated, database, data entry still remains a heavy chore, and may, at worse case, still take thirty hours to enter a day's races. One way to get around this problem a little bit, is to narrow the field of horses which we are interested in handicapping.

For example, if we concentrated our handicapping efforts on horses in the claiming classes $3000 to $5000, we would only have to cover about half of the horses which run frequently at Suffolk Downs. This would reduce our database to half it's full size without significantly reducing the number of copied past performances. Our redundancy rate would still be about seven out of eight races, because the past performances for the horses which we would be entering would be primarily races in the $3000 to $5000 dollar claiming classes (with a few exceptions of horses moving up and down in class into and out of this range.) Thus, we would only have to enter half of the total number of horses which frequent Suffolk Downs to have a reasonably

complete database for $3000 - $5000 claimers. The fact that our database would be complete for this range means that all of the data-entry time saving features would still work, and that it would take less time (about half as much) to initiate the half database, and to maintain it. While this also means that the number of races which you could handicap would also be cut in two (you'd only have data for claiming races of class $3000 to $5000), this probably would not be a serious problem to the handicapper. If the issue was making a certain dollar amount of profit per day, a handicapper would just increase the amount he would bet, while just betting on those races which fall within the $3000 - $5000 claiming range.

My tests on the WHINNY input system have verified that it is possible to reduce the time needed for data entry to something reasonable. Unfortunately, this is only possible after one establishes a somewhat complete database of horse entries and past performances, which the system can use for it's time saving measures. Fortunately, it is not necessary to establish an absolutely complete database; a relative database for a certain type and class of race will also establish the needed redundancy in the entries to enable the time-saving features. While this limitation would reduce the number of "bettable" races, it would also enable the not-so-dedicated handicapper to use the WHINNY system without spending excessive time establishing a totally complete database.

Analysis Testing

The second aspect of the WHINNY system that required testing was the success of the handicapping analyses. As I mentioned previously, I had programmed four main handicapping methods: *Dr. Quirin's Sprint Formula*, *Dr. Quirin's Route Formula*, *Ainslie's Rules of Thumb*, and *Quirin and Ainslie Combined Methods.* For each race I entered, I tested three of the methods; only the one of Quirin's formula was used, depending on whether the race was a sprint or a route.

One of the overwhelming problems of testing this part of the system was the time needed to enter a race to be handicapped. While I have shown in my testing that it is possible to update a full database in only ~10 minutes per race, this unfortunately is only true if one has a full database to begin with. At the time of this writing, my database is only partially begun, and it is usually the case that I had to enter all of the horses in a race from scratch. This meant that before I could analyze a race, I had to first spend two to four hours entering the horses participating in the race. Due to this time factor, the number of races on which I tested my analyses is too small to draw very accurate conclusions from. On a rough approximation, however, it appears that the WHINNY analysis system is performing well enough to turn some substantial profit.

I was able to enter three complete races which I used to test my analyses. In all three cases, the method *Ainslie's Rules of Thumb* was not able to find any spot plays for me to bet on. This indicated that the averaging method used in Quirin's analyses would be adequate for analyzing the races.

Two of my three races required *Dr. Quirin's Sprint Formula*, while the third race was a route, thus employing *Dr. Quirin's Route Formula*. In the results of all three of the analyses, one horse was clearly superior to the rest of the field. These were actual races, for which real results were obtainable. For two of the races, the WHINNY system picked the first horse to cross the finish line. In one of the races, this "best" horse was disqualified for interfering with another horse during the race, and placed second (sigh). In the third case, the horse picked the WHINNY system came in fourth, but it was only one length behind the leader. (It was a very close race.)

Thus out of three races, the WHINNY system was able to pick the winners of two of the races. If a bettor had bet $2 on the horse picked by the WHINNY system in each case, he would have a total loss of $1.20, due to the disqualification of one of the winning horses. This would be a loss of 20 cents on every dollar bet (the equivalence of the track take of 20 cents which comes out of every dollar bet, i.e. almost breaking

even). If the horse in the second race had not been disqualified, however, the bettor would have made a profit of $2.80, a profit of 46 cents on each dollar bet. As these figures indicate, it is not clear at this time, whether the WHINNY system is actually capable of turning a profit, in the long run. More testing is needed to determine this.

I have included as an example the analysis of one of the three races. On the following pages, I will show the horses entries for the horses entered in the race, the race description as given to the WHINNY system, and then the results returned by the WHINNY system, and the actual results of the race, itself.

The race I am enclosing was the first race on 20 March 85, at Suffolk Downs. The horse descriptions are on the first three pages, followed by a page showing the layout of the race, the results of the WHINNY analysis, and the payoffs for the actual results. For this race, the analysis used was *Dr. Quirin's Sprint Formula.* Figures 9, 10, 11 and 12

Whinny

Charbonnel — B h 6, by Roberto -- Chocolate Beau, by Beau Max
Br.-- Hanes Mrs J W (ky) Own.-- Garren M M Tr.-- Verdillo Frank
Lifetime $43,045 1985 3 0 1 0 $700
56 4 5 4 1984 10 1 0 1 $8,770
Turf 20 1 3 3 $9,505

Date	R#	Trk	Cnd	Dist	1st	2nd	Finsh	Type	Amnt	Restr	PP	Brk	1st	2nd	3rd	Fin	Jockey	Eqp	Wt	Odds	SR-TV	Horses(1st 2nd 3rd)weights and lengths	Comment	#Horses
6Mar85	- 8	Suf	fst	6f	:22⅖	:47⅗	1:14⅗	Clm	4250		8	1	6^7¼	2^2	1^nk	2^2¼	Klein M D	b	114	12.90	60-34	FinalAce 119^2¼ Charbonnel 114^8¼ BigStitch 119^2	Hung	8
2Feb85	-10	Suf	fst	6f	:22⅖	:46	1:13⅖	Clm	7500		9	7	11^13	12^10	12^12	11^10¼	Klein M D	b	114	20.70	59-32	PresidentialMay 119^nk PreciousPower 109^2¼ DisplayBid 112^2¾	Outrun	12
15Feb85	- 7	Suf	gd	6f	:23	:47⅖	1:14	Clm	7500		8	8	11^14¼	9^10	9^8¼	7^4¼	Klein M D	b	114	37.00	55-32	MountainBarb 113¾ Noxexplode 115^1¾ DevilFish 114¾	Wide	12
13Jun84	- 6	Suf	fst	6f	:22⅖	:46	1:10⅖	Clm	7500	3†	5	3	5^5	6^3¾	6^3¾	6^5	Giacomo A R	b	111	7.00	70-10	EnergyStrand 110^3¾ BudasTime 119^4¾ Innocuous 115^2	No Threat	7
27Jun84	- 7	Suf	fst	6f	:22⅖	:45⅖	1:12	Clm	10000	3†	3	3	4^2	5^2¼	5^2¼	5^4¾	Di Mauro S D	b	105	20.90	76-26	RedSultan 117^2¾ HouseOdds 115^nk SealedConquest 117^nk	No Factor	6
16Jun84	- 8	Suf	fst	6f	:23	:46⅖	1:11⅗	Clm	16500	3†	8	2	7^4¾	7^4¾	7^4¾	6^2¼¾	Donahue G W	b	111	46.20	71-25	LittleMoreTime 114^2¾ FoolishMove 114^nk AcuteSymptom 119^5	Outrun	8
1Jun84	- 2	Bel	fst	6f	:22⅖	:44⅖	1:10⅖	Clm	14000		6	9	9^7	8^10	7^12	7^12	Vega H	b	110	23.90	79-20	LittleMatch 115^1¾ Rasselas 122^nk BrightSearch 117^1¾	Outrun	9
17May84	- 2	Bel	fst	6f	:23⅖	:46⅖	1:10⅖	Clm	16000		2	8	9^nk	8^10	7^11	6^4¾	Smith A Jr	b	117	10.20	76-17	Kotter 105^1¾ Rasselas 113^2¾ BigHughes 119^1¾	Outrun	9

Whinny

Internie Look — Dk B c 4, by Wrinkle Dinkie -- Jo Jan, by Sails Pride
Br.-- Siculiana Farm (ont C) Own.-- Prince J E Tr.-- Prince John E
Lifetime $37,323 1985 6 1 1 1 $3,130
31 6 9 3 1984 14 1 3 2 $11,777
Turf 0 0 0 0 $0

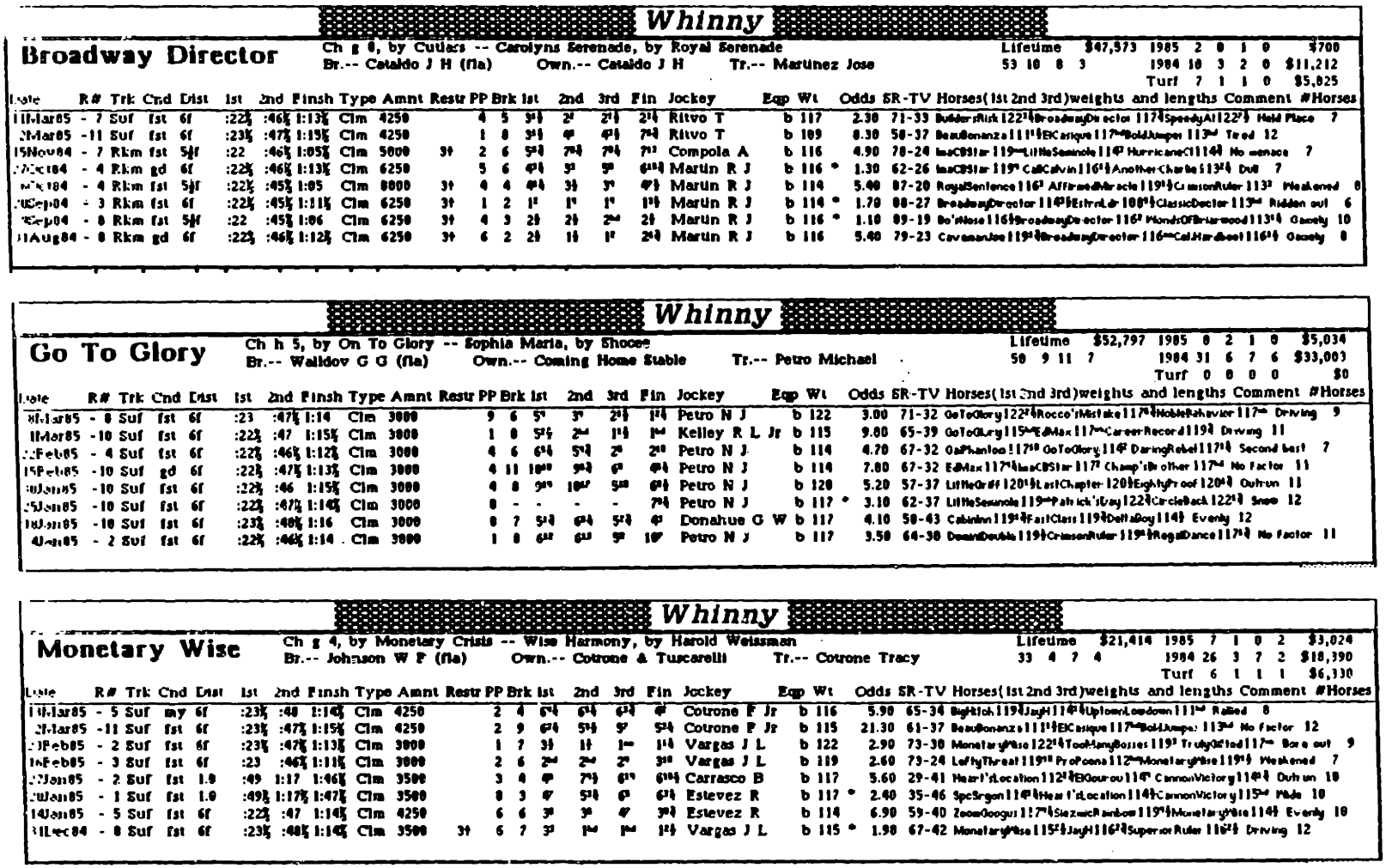
Date	R#	Trk	Cnd	Dist	1st	2nd	Finsh	Type	Amnt	Restr	PP	Brk	1st	2nd	3rd	Fin	Jockey	Eqp	Wt	Odds	SR-TV	Horses(1st 2nd 3rd)weights and lengths	Comment	#Horses
11Mar85	- 5	Suf	fst	6f	:23⅖	:47⅖	1:14⅖	Clm	3000		9	2	2^5	2^2	2^2½	4^3¼	Elliott S		119	* 1.50	66-33	GotDownOnIt 122^2 HeavyMetal 119^2¾ MoeToMo 119^nk	Weakened	11
25Feb85	- 9	Suf	fst	6f	:23	:48⅖	1:15⅖	Clm	4250		4	1	3^2	2^nk	1^1½	2^1¾	Santiago M D		111	3.30	61-40	FinalAce 114^6¾ InternetLook 111^1¾ Matt'sChoice 116^10	Weakened	5
2Feb85	- 2	Suf	fst	6f	:23⅖	:47⅖	1:13⅖	Clm	3000		2	4	1^2	1^1½	1^2	1^1¾	Elliott S		122	* 1.30	72-32	InternetLook 122^2¾ TrulyOffled 117^4¾ SteelPier 122^nk	Driving	10
20Jan85	- 3	Suf	fst	6f	:23⅖	:47⅖	1:14⅖	Clm	4250		5	1	2^nk	2^nk	2^2½	3^3¾	Elliott S		114	5.50	60-37	BeauBananza 114^6¾ Matt'sChoice 116^1¾ InternetLook 114^2	Weakened	7
11Jan85	- 2	Suf	fst	6f	:22⅖	:46⅖	1:14	Clm	4250		4	2	2^nk	2^2	3^3½	7^10¾	Elliott S		114	7.40	60-37	OhApproach 113^2¾ SuzmicRainbow 116^1 HaliveBengal 116^nk	Tired	10
11Jan85	- 8	Suf	fst	6f	:23	:48	1:15⅖	Clm	7500		7	2	3^2½	8^11	11^12	11^12¾	Elliott S		114	22.00	42-46	GeeJo 117^1¾ IrishKing 114^6¾ Karino 114^4¾	Brief speed	11
26Dec84	- 9	Suf	fst	6f	:22⅖	:47⅖	1:15⅖	Clm	10000		5	3	2^½	4^3¾	4^3½	8^11¾	Riera R Jr		115	51.20	51-39	PresidentialMay 114^6¾ AllKeyBoy 114^2 JockyBo 117^nk	Tired	9
9Sep84	- 9	Det	sly	6f	:22⅖	:46⅖	1:13⅖	Clm	5000	3†	6	4	1^½	2^2	3^3½	7^4¾	Allen R D Jr		115	5.90	68-29	Governor88 120^2¾ EarlyEscape 115^nk OutRunner 115^nk	Tired	10

Whinny

Irish Poplar — Ch h 6, by Irish Castle -- Tulip Poplar, by Some Chance
Br.-- King H G III (ky) Own.-- Souto Gerald Tr.-- Souto Gerald
Lifetime $50,449 1985 6 0 0 0 $230
73 11 9 8 1984 19 4 1 3 $10,502
Turf 6 1 0 0 $10,000

Date	R#	Trk	Cnd	Dist	1st	2nd	Finsh	Type	Amnt	Restr	PP	Brk	1st	2nd	3rd	Fin	Jockey	Eqp	Wt	Odds	SR-TV	Horses(1st 2nd 3rd)weights and lengths	Comment	#Horses
8Mar85	- 7	Suf	gd	1^70	:47⅖	1:14⅖	1:46⅖	Clm	3000		4	1	1^2	2^2	6^11	7^12	Caraccio S	b	119	14.60	47-32	TangoVictor 117^10 Norm 119^nk CannonVictory 114^6¾	Stopped	8
1Mar85	-10	Suf	fst	6f	:22⅖	:47	1:15⅖	Clm	3000		9	6	7^4	6^3½	5^2½	8^6	Vargas J L	b	119	50.50	59-39	GoToGlory 115^nk EdMax 117^nk CareerRecord 119^3	Steadied	11
19Feb85	- 2	Suf	fst	6f	:23⅖	:47⅖	1:14⅖	Clm	3200		3	7	6^3¾	6^3	5^3½	7^12	Vargas J L	b	122	3.80	52-35	SpursAndFeathers 122^2¾ Moongraph 115^5 Ransera 114^8	Outrun	7
28Jan85	- 4	Suf	fst	1^70	:49⅖	1:16⅖	1:50⅖	Clm	3000		6	4	3^1½	3^2	4^12	5^14	Ritvo T	b	117	13.70	32-46	Norm 119^1¾ OneWinner 122^10 RosnatJA 119^2	Early pool	9
14Jan85	- 3	Suf	fst	6f	:23	:47⅖	1:16⅖	Clm	3200		8	6	4^2¾	4^2	5^2½	6^11¾	Cotrone F Jr	b	122	3.10	46-40	LittleSeminole 122^3¾ NorthernPearl 122^2¾ CrashKeough 122^1¾	Tired	8
3Jan85	- 5	Suf	fst	6f	:23	:47⅖	1:14⅖	Clm	3200		4	8	8^7¾	7^5¾	7^5¾	6^12¾	Vargas J L	b	122	* 0.90	57-34	BillyBeGone 122^4¾ NorthernPearl 119^1¾ Who'sUgly 114^6¾	Steadied	8
26Dec84	- 2	Suf	fst	5f	:23⅖	:47⅖	1:15	Clm	3200	3†	3	9	6^3½	4^3½	1^nk	1^nk	Vargas J L	b	122	* 1.60	66-39	IrishPoplar 122^3¾ MakingRipples 120^1¾ OhApproach 122^2nd	Driving	9
17Dec84	- 3	Suf	fst	6f	:23⅖	:47⅖	1:15⅖	Clm	3000	3†	2	4	1^1½	1^nk	1^1¾	1^1½	Vargas J L	b	120	2.10	62-44	IrishPoplar 120^1¾ PunchoutPrince 120^2 Fierce 120^6¾	Handily	10

Figure 9: Horses in sample race



Whinny

Broadway Director — Ch g 6, by Cutlass -- Carolyns Serenade, by Royal Serenade
Br.-- Cataldo J H (fla) Own.-- Cataldo J H Tr.-- Martinez Jose

	Lifetime	1985	1984	Turf
	$47,573	2 0 1 0 $700	10 3 2 0 $11,212	7 1 1 0 $5,825
	53 10 8 3			

Date	R#	Trk	Cnd	Dist	1st	2nd	Finsh	Type	Amnt	Restr	PP	Brk	1st	2nd	3rd	Fin	Jockey	Eqp	Wt	Odds	SR-TV	Horses(1st 2nd 3rd)weights and lengths	Comment	#Horses
11Mar85	- 7	Suf	fst	6f	:22⅖	:46⅖	1:13⅖	Clm	4250		4	5	3hd	2d	2hd	2nk	Ritvo T	b	117	2.30	71-33	BuddersRisk 122nk BroadwayDirector 117hd SpeedyAl 122nk	Held Place	7
2Mar85	-11	Suf	fst	6f	:23⅖	:47⅖	1:15⅖	Clm	4250		1	8	3hd	4d	4hd	7nk	Ritvo T	b	109	8.30	58-37	BeauBonanza 111nk ElCarique 117nk BoldJumper 113nk Tired		12
15Nov84	- 7	Rkm	fst	5½f	:22	:46⅖	1:05⅖	Clm	5000	3†	2	6	5hd	7nk	7nk	7nk	Compola A	b	116	4.90	78-24	IraCBStar 119nk LilHeSeminole 114d HurricaneC 114d	No menace	7
2Oct84	- 4	Rkm	gd	6f	:22⅖	:46⅖	1:13⅖	Clm	6250		5	6	4hd	3d	5d	6nk	Martin R J	b	116 *	1.30	62-26	IraCBStar 119d CallCalvin 116nk AnotherCharlie 113nk	Duel	7
4Sep84	- 4	Rkm	fst	5½f	:22⅖	:45⅖	1:05	Clm	8000	3†	4	4	4hd	3nk	3d	4hd	Martin R J	b	114	5.40	87-20	RoyalSentence 116d AffirmedMiracle 119nk CrimsonRuler 113d	Weakened	8
20Sep84	- 3	Rkm	fst	6f	:22⅖	:45⅖	1:11⅖	Clm	6250	3†	1	2	1d	1d	1d	1nk	Martin R J	b	114 *	1.70	88-27	BroadwayDirector 114nk EthrnLdr 108nk ClassicDoctor 113hd	Ridden out	6
5Sep84	- 8	Rkm	fst	5½f	:22	:45⅖	1:06	Clm	6250	3†	4	3	2nk	2nk	2hd	2nk	Martin R J	b	116 *	1.10	89-19	Bo'nHose 116nk BroadwayDirector 116d MondsOfBirnamwood 113nk	Gamely	10
31Aug84	- 8	Rkm	gd	6f	:22⅖	:46⅖	1:12⅖	Clm	6250	3†	6	2	2nk	1hd	1d	2nk	Martin R J	b	116	5.40	79-23	CavemanJoe 119nk BroadwayDirector 116nk CalHardfoot 116nk	Gamely	8

Whinny

Go To Glory — Ch h 5, by On To Glory -- Sophia Maria, by Shocee
Br.-- Walldov G G (fla) Own.-- Coming Home Stable Tr.-- Petro Michael

	Lifetime	1985	1984	Turf
	$52,797	0 2 1 0 $5,034	31 6 7 6 $33,003	0 0 0 0 $0
	50 9 11 7			

Date	R#	Trk	Cnd	Dist	1st	2nd	Finsh	Type	Amnt	Restr	PP	Brk	1st	2nd	3rd	Fin	Jockey	Eqp	Wt	Odds	SR-TV	Horses(1st 2nd 3rd)weights and lengths	Comment	#Horses
8Mar85	- 8	Suf	fst	6f	:23	:47⅖	1:14	Clm	3000		9	6	5d	3d	2hd	1nk	Petro N J	b	122	3.00	71-32	GoToGlory 122nk Rocco'sMistake 117hd NobleBehavior 117nk	Driving	9
11Mar85	-10	Suf	fst	6f	:22⅖	:47	1:15⅖	Clm	3000		1	8	5hd	2nk	1hd	1nk	Kelley R L Jr	b	115	9.00	65-39	GoToGlory 115nk EdMax 117nk CareerRecord 119d	Driving	11
2Feb85	- 4	Suf	fst	6f	:22⅖	:46⅖	1:12⅖	Clm	3000		4	6	6hd	5hd	2d	2nk	Petro N J	b	114	4.70	67-32	OaPhantom 117nk GoToGlory 114d DaringRebel 117hd	Second best	7
15Feb85	-10	Suf	gd	6f	:22⅖	:47⅖	1:13⅖	Clm	3000		4	11	10nk	9nk	6d	4hd	Petro N J	b	114	7.80	67-32	EdMax 117hd IraCBStar 117d Champ'sBrother 117nk	No Factor	11
30Jan85	-10	Suf	fst	6f	:22⅖	:46	1:15⅖	Clm	3000		4	8	9nk	10nk	5nk	6hd	Petro N J	b	120	5.20	57-37	LilHeGrr 120hd LastChapter 120nk EghtyProof 120nk	Outrun	11
25Jan85	-10	Suf	fst	6f	:22⅖	:47⅖	1:14⅖	Clm	3000		8	-	-	-	-	7nk	Petro N J	b	117 *	3.10	62-37	LilHeSeminole 119nk Patrick'sDay 122d CircleBack 122nk	Snee	12
18Jan85	-10	Suf	fst	6f	:23⅖	:48⅖	1:16	Clm	3000		8	7	5hd	6hd	5hd	6d	Donahue G W	b	117	4.10	50-43	Cabinlnn 119hd FastClass 119nk DeltaBoy 114d	Evenly	12
4Jan85	- 2	Suf	fst	6f	:22⅖	:46⅖	1:14	Clm	3000		1	8	6nk	6nk	5d	10d	Petro N J	b	117	3.50	64-30	DeanDouble 119d CrimsonRuler 119nk RegalDance 117hd	No Factor	11

Whinny

Monetary Wise — Ch g 4, by Monetary Crisis -- Wise Harmony, by Harold Weissman
Br.-- Johnson W F (fla) Own.-- Cotrone & Tuscarelli Tr.-- Cotrone Tracy

	Lifetime	1985	1984	Turf
	$21,414	7 1 0 2 $3,024	26 3 7 2 $18,390	6 1 1 1 $6,330
	33 4 7 4			

Date	R#	Trk	Cnd	Dist	1st	2nd	Finsh	Type	Amnt	Restr	PP	Brk	1st	2nd	3rd	Fin	Jockey	Eqp	Wt	Odds	SR-TV	Horses(1st 2nd 3rd)weights and lengths	Comment	#Horses
11Mar85	- 5	Suf	my	6f	:23⅖	:48	1:14⅖	Clm	4250		2	4	6nk	6nk	6nk	6d	Cotrone F Jr	b	116	5.90	65-34	BigRich 119nk JayH 114nk UptownLowdown 111hd	Rallied	8
2Mar85	-11	Suf	fst	6f	:23⅖	:47⅖	1:15⅖	Clm	4250		2	9	6nk	5hd	5d	5nk	Cotrone F Jr	b	115	21.30	61-37	BeauBonanza 111nk ElCarique 117nk BoldJumper 113hd	No Factor	12
20Feb85	- 2	Suf	fst	6f	:23⅖	:47⅖	1:13⅖	Clm	3000		1	7	3d	1d	1hd	1nk	Vargas J L	b	122	2.90	73-30	MonetaryWise 122nk TooManyBosses 119d TrulyDotted 117hd	Bore out	9
16Feb85	- 3	Suf	fst	6f	:23	:46⅖	1:11⅖	Clm	3000		2	6	2hd	2hd	2d	3nk	Vargas J L	b	119	2.60	73-24	LeftyThreat 119nk ProPcona 112hd MonetaryWise 119d	Weakened	7
27Jan85	- 2	Suf	fst	1.0	:49	1:17	1:46⅖	Clm	3500		3	4	4d	7nk	6nk	6nk	Carrasco B	b	117	5.60	29-41	Heart'sLocation 112nk ElGourou 114d CannonVictory 114nk	Outrun	10
12Jan85	- 1	Suf	fst	1.0	:49⅖	1:17⅖	1:47⅖	Clm	3500		8	3	6d	5nk	6d	6nk	Estevez R	b	117 *	2.40	35-46	SpcSrgn 114nk Heart'sLocation 114nk CannonVictory 115hd	Wide	10
4Jan85	- 5	Suf	fst	6f	:22⅖	:47	1:14⅖	Clm	4250		6	6	3d	3d	4d	3hd	Estevez R	b	114	6.90	59-40	ZeusGoogul 117nk SiezwicRainbow 119nk MonetaryWise 114nk	Evenly	10
31Dec84	- 8	Suf	fst	6f	:23⅖	:48⅖	1:14⅖	Clm	3500	3†	6	7	5d	1hd	1hd	1nk	Vargas J L	b	115 *	1.90	67-42	MonetaryWise 115nk JayH 116nk SuperiorRuler 116nk	Driving	12

Figure 10: Horses in sample race

Whinny

Pro Poona — B g 5, by Thumpro -- Fly Poona, by Exalted Ruilah
Br.-- Horsefly Farms (colo) Own.-- Pace A Tr.-- Pace A

Lifetime $24,350 1985 8 2 1 1 $5,144
53 7 9 5 1984 28 1 3 2 $7,277
Turf 0 0 0 0 $0

Date	R#	Trk	Cnd	Dist	1st	2nd	Finsh	Type	Amnt	Restr	PP	Brk	1st	2nd	3rd	Fin	Jockey	Eqp	Wt	Odds	SR-TV	Horses(1st 2nd 3rd)weights and lengths	Comment	#Horses
13Mar85	-10	Suf	fst	1.0	:48⅘	1:15⅘	1:44⅘	Clm	3200		3	3	2ⁿᵏ	1¹½	1¹½	1ⁿᵉ	Ritvo T	b	109	4.30	55-30	ProPoona 109ⁿᵉ CambridgeAce 111¹½ Matt'sChoice 114⁴	Driving	10
2Mar85	- 1	Suf	fst	6f	:23⅘	:48⅘	1:15⅘	Clm	3000		8	7	7⁸¼	6⁵¾	5³¾	3¾	Ritvo T	b	114	25.30	61-37	JayH122¾ ReasonToHope 122ⁿᵏ ProPoona 114¹¼	Rallied	9
25Feb85	- 2	Suf	fst	6f	:23⅘	:49	1:17	Clm	3000		2	4	4²½	4⁴	1¹½	1¹½	Ritvo T	b	109	1 30	56-40	ProPoona 109¹¼ NapoleonHill 122ⁿᵏ HoldAsoon 117¹¼	Driving	11
16Feb85	- 3	Suf	fst	6f	:23	:46⅘	1:11⅘	Clm	3000		3	2	3ⁿᵏ	3ⁿᵏ	3⁵	2¹⁰	Ritvo T	b	112	10.80	73-24	LoftyThreat 119¹⁰ ProPoona 112ⁿᵏ MonetaryWise 119¹¼	Second best	7
11Feb85	-10	Suf	gd	6f	:23	:47	1:13⅘	Clm	3000		8	10	8²¼	6⁸	4⁸¼	4⁸¼	Ritvo T	b	109	23.2(	51-30	MedievalManor 117¹¾ Zunami 114⁶¼ TheMangler 114⁸¼	No Factor	12
5Feb85	- 2	Suf	fst	6f	:23	:47⅘	1:15	Clm	3000		2	7	4⁶	7⁸¼	6³¾	6³¾	Ritvo T	b	109	22.30	60-40	ReasonToHope 122ⁿᵏ OscarWinner 114⁸½ BurntGypsum 117¹¾	Tired	10
26Jan85	- 2	Suf	fst	6f	:23½	:46⅘	1:16⅘	Clm	3000		5	9	6⁸½	7⁸¾	6⁷¾	7⁸	Baez R	b	115	10.30	53-38	PrestoChango 117ⁿᵏ Legrand 114ⁿᵉ RideOnMax 119⁴¾	Outrun	11
18Jan85	-10	Suf	fst	6f	:23⅘	:46⅘	1:16	Clm	3000		4	5	2ᵈ	2ⁿᵏ	4ⁿᵉ	6⁶¾	Santiago M D	b	112	36.10	56-43	Cabildon 119¹¾ FastClass 119¾ DattaBoy 114¾	Tired	12

Whinny

Bachelor Dinner — Ch g 10, by Quadrangle -- Bethrothed, by Bold Ruler
Br.-- Janney Mr Mrs S E Jr (md) Own.-- Petrillo Anthony Tr.-- Petrillo Anthony 147 18 15 18

Lifetime $87,166 1985 3 0 0 0 $315
1984 31 2 0 6 $8,578
Turf 2 0 0 0 $0

Date	R#	Trk	Cnd	Dist	1st	2nd	Finsh	Type	Amnt	Restr	PP	Brk	1st	2nd	3rd	Fin	Jockey	Eqp	Wt	Odds	SR-TV	Horses(1st 2nd 3rd)weights and lengths	Comment	#Horses
2Jan85	- 2	Suf	fst	6f	:22⅘	:45⅘	1:14	Clm	4250		5	9	8⁸¼	7¹⁰	7¹³	5⁸¼	Paulino G A	b	119	36.20	62-37	OldApproach 113¹¾ SiezmicRainbow 116³ NativeBengal 116ⁿᵏ	Outrun	10
14Jan85	- 5	Suf	fst	6f	:22⅘	:47	1:14⅘	Clm	4250		4	7	5⁵	5¹¹	6¹²	5¹⁰¾	Paulino G A	b	117	35.30	58-40	ZoomGeogur 117¹¾ SiezmicRainbow 119¹¾ MonetaryWise 114¾	Evenly	10
7Jan85	- 7	Suf	fst	6f	:22⅘	:46⅘	1:12	Clm	4250		5	5	7⁸¾	6⁶	6¹⁰	5¹⁰¾	Gonzalez I O	b	112	19.20	70-28	Stan'sChoice 117³ FirstProz 112¹¾ Matt'sChoice 122¾	Wide	8
2Dec84	- 3	Suf	fst	6f	:22⅘	:47⅘	1:14⅘	Clm	4250	3†	8	3	5ⁿᵏ	5⁵	5¹¹	5ⁿᵏ	Gonzalez I O	b	114	22.90	58-39	RegalRocky 114⁴ Norm 112² CareerYankee 109⁸¾	No threat	8
17Dec84	- 3	Suf	fst	6f	:22⅘	:46⅘	1:13⅘	Clm	4250	3†	5	5	5⁵	6⁶	7⁸¾	5¹³¾	Gonzalez I O	b	114	58.10	59-30	Decatur 114⁴ SnappyDelivery 119¹¾ MountainBarb 119¹	No Factor	8
1Dec84	- 5	Suf	fst	6f	:23	:46⅘	1:16⅘	Clm	4250	3†	9	6	6⁷¾	5⁷¾	6⁷	7¹⁰	Gonzalez I O	b	112	35.10	48-52	NativeBengal 114¹¾ Vie d'Fantastic 122ⁿᵏ LilHeMoreTime 117¾	Outrun	10
24Nov84	- 4	Suf	fst	6f	:23⅘	:46⅘	1:16⅘	Clm	4250	3†	4	3	4⁶	5¹²	5¹⁰	5¹²	Gonzalez I O	b	114	6.60	41-45	Rexexplode 114⁴ NativeBengal 114¾ Norm 114ⁿᵉ	Wide	5
14Nov84	- 7	Rkm	fst	6f	:22⅘	:46⅘	1:13⅘	Clm	4000	3†	5	7	5³	3¾	4⁶¾	3⁴	Gonzalez I O	b	116	41.40	77-26	PolishFalcon 113ⁿᵏ LilHeMoreTime 113¹ BachelorDinner 116³	Brushed	10

Figure 11: Horses in sample race.

Figure 12: Description of Race, and Results of Analysis

Figure 13: Historgrams of Analysis Results for Three Races

It is somewhat informative to look at the predicted results of the other two races which I handicapped, along with the official results of the races. I have constructed histograms of all three races, which you can find as Figure 13. This information is useful in gaining a feeling for the scale and distribution of the point values returned by the Quirin tests. (The method *Ainslie's Rules of Thumb* was not able to locate an spot plays in any of these three races.)

Another Look at Handicapping Models

The WHINNY system, as it stands, only implements a model which treats the handicapping factors as if they were linear in nature. Yet, we have argued, in Chapter 5, that an attack more along the lines of an expert system would be a better model. There are many ways in which an expert system model could be designed. For example, the non-linear aspects in Quirin's formulae could be modified and used to direct the proportions assigned to the other tests. Or an expert system could be devised, in which the flow from test to test would be directly influenced by the results from earlier tests. It would be easy enough to construct such a system, given the current state of knowledge about handicapping, along with the ability to use the WHINNY system. At any point, if we were unsure about where the analysis should branch to, we could conduct mass testing of the WHINNY database to determine the effectiveness of the various tests under consideration, given a set of previously defined conditions (conditions which would describe the results of earlier tests in the system "tree" up to this point). It is not necessary to attempt to construct this expert system from scratch, however. We can simplify the problem down to several major tests: 1) How fast can the horse run in top condition, 2) Will the horse be in top condition for this race, and 3) How will other exterior factors such as jockey skill or speed horses influence our horse? For each of these major test categories, there are several tests which may be used to determine the answers. Our expert system could select the most successful test for this situation, based on its knowledge of the situation, and factors such as class and distance. I will go into this model in a little more detail later on. First I would like to step you through an analysis of the race using a very primitive type of expert system. This system, though inferior to a truly "expert" system, is very similiar to the train of thought used by many so-called "expert" handicappers, as they pour over their *Daily Racing Form* on the night before the races. It would also be rather easy to program in the WHINNY system, do to it's simplicity.

I will illustrate this simple expert system by using the horses and the race on the previous pages as examples for one type of exploratory process. In this system, tests are arranged in order of priority, i.e. those tests considered to be more important are used first, with later tests only used if earlier tests do not eliminate all but one horse. Horses which do not meet the "minimum" competency level for each test are eliminated from the competition. The remaining "surviving" horses are passed on to the next test until only one "best" horse remains. This is a very simple example of an expert system, and definitely not the best model. Rather than eliminating horses after they fail just one test, they should instead be checked for "redeeming" features or possible "interference" value. (By "interference" I mean that by their mere presence in the race, they may contribute some factor which will influence the rest of the horses in the rest. For example, a "speed" horse which sets a very fast pace for the first half mile would have an effect on other "speed" horses in the race.) For such reasons it would make better sense to try some different tests or branches of tests on the failing horses. But, for the sake of simplicity, in our example, horses failing any one test shall be discarded from the analysis.

Start with one handicapping factor. Let's pick "class" as our starting place. There are three possible things which could happen in class. A horse could be moving up in class, staying in the same class, or moving down in class. Horses moving up in class should look like they have the ability to tackle better quality horses. Horses staying in a class should look like they can at least do satisfactorily in that class. Horses dropping in class should have either shown some life in their last race, a sign that they are still capable of running despite the fact that their trainer thinks that they are not capable of handling that higher class. If any of these factors are not present and an excuse is not found, the horse's chances of doing well in this race are severely limited. An excuse would be if the last race was at the wrong distance (i.e. a sprint when today's race is a route, and the horse has shown routing ability, or visa versa, the last races was a route,

and today's race is a sprint.) As mentioned above, horses which do not meet the criteria of this "class" test shall be eliminated.

Lets look at our horses. Those moving up in class include: Internie Look, Irish Popular, Go to Glory, and Pro Poona. Of these four horses, only two, Go to Glory, and Pro Poona showed that they have the ability to move up in class (Pro Poona looks weak (she only won her last race by a neck), but let's keep her for now). Those staying in the $4250 class included: Charbonnel, Broadway Director, Monetary Wise, and Bachelor Dinner. Of these four, only three had reasonable last races. Bachelor Dinner failed miserably, coming in 5th by 8 and 1/2 lengths. He gets discarded from our analysis. The other three horses stay. There are no horses moving down in class. So, in our analysis, we are down to five horses from the eight initial ones. Our system moves on to check the speed ratings for these five horses.

The reasoning behind testing speed ratings is as follows. Speed ratings determine which horse has the greatest potential for running fastest. Other conditions, such as current condition or jockey will determine whether the horse runs as fast as he can today. Appropriate use of body language, discussed in the next chapter, is also a good measure of determining a horse's condition on any day. But, for this next step, lets just try to find those horses which have the speed, on a good day, to win the race. We shall, in this analysis, eliminate all horses whose speed rating is less than three points below the speed rating of the fastest horse. Note that this is just an arbitrary cutoff point. Maybe two point or four points would be better. This sort of thing could be tested for, on a sufficiently large database to determine where a reasonable cutoff should be. For now, however, let's assume that three points is reasonable.

Fortunately for us, the WHINNY system includes an primitive test *Speed Rating Test* which returns as results the speed ratings for each horse. Our expert system applies this to the race to determine the speed ratings for our horses under consideration. The results of this test are included on the next page. From the five horses still in our

analysis, Pro Poona and Monetary Wise are eliminated. Their speed ratings are not within three points of the highest speed rating. Thus, our analysis has eliminated five of the original eight horses. The remaining horses are: Broadway Director, Go to Glory, and Charbonnel. The next thing we will looks for, in these three horses, is an increase in weight, beyond that which these horses have carried in the past.

There is no doubt among handicappers that weight slows a horse down. Track handicappers assign weights for just this purpose, to even out a field of horses to make the race more fair. But sometimes, a drastic increase in weight can eliminate a potential candidate. A good rule of thumb is to check each horse's past performances, and determine if the horse has ever carried before the weight which he will be carrying today, and if so, did the weight have an impact on the horse's performance.

Let's looks at our horses. Broadway Director is carrying 115 pounds in this race, yet he carried 117 in the last race successfully. No trouble there. Go to Glory is carrying 117 pounds in this race, after carrying 120 in the previous race. No trouble here either. Looking at Charbonnel, however, we find the makings of a problem. Charbonnel has been assigned the high weight of 122 pounds, a six pound increase over the 114 weight carried in his last race. Looking at Charbonnel's past performances, we find that the horse has never carried such a high weight before. Not good. On this basis, we eliminate Charbonnel from our analysis.

Are either of our two final contenders "speed" horses? In other words, do they like to get out early and lead the pack? If so, is there the possibility of a speed duel? In a speed duel, involving two or more speed horses, it is often the case that the speed horses expend all of their energy early in the race, leaving the winning to a horse which typically runs "off the pace" (in the middle of the pack), and then overtakes the leaders on or just after the final turn. If either of our two horses is a speed horse, and if there is the chance of a speed duel with any of the other horses in the race (even the ones we have eliminated), this could be the deciding factor in choosing the "best" horse.

Looking at our two horses, Broadway Director, and Go to Glory, we discover that neither of these horses is a "speed" horse. If there is a speed duel in this race, these horses will probably not be involved, as both are slow breakers, as indicated by their past performances. This step of the analysis has not eliminated any horses. And we still have not picked a winner.

Our system continues to picks it's way through several more tests, trying to eliminate one of these two horses. A good test to use here is jockey quality. The two horses are so close in quality that if one of them has a better jockey, that would be good enough to win the race. We find, however, by checking the jockeys, that they both have had successful races on their horses, and that they both are good jockeys. No eliminations here.

How about changes in distance? Are both of these horses accustomed to today's sprint distance of six furlongs, or is one of these horses typically a router. A quick check of their past performances indicates that both horses are sprinters. So much for that.

Another good test is to check for layoffs. A horse typically does it's best on its third or fourth race after a layoff. This was proven by the other analyses done by Dr. Quirin, when he checked each possible handicapping factor for its effectiveness. [Qurn79] While he did not use this test in either of his formulae, I have found from my own handicapping experience that it is very effective, especially when no other distinguishing factors can be found. Well, let's looks at our horses.

Go to Glory has been running constantly throughout the winter. No layoff here. But Broadway Director has only had two races since a four month layoff. Hmmm. And his last race was significantly better than the one before that. Today is his third race after a layoff. According to Quirin, there is a good probability that he will do better in today's race than in his last one. And that should be good enough to beat Go to Glory. On this final test, we eliminate Go to Glory as winner. He is, however, easily the second best horse in the field, as it took many tests before a deciding difference could

be found between him and Broadway Director.

This has been one example of how an expert system could be used to step through and handicap a race. There are many other tests which I could have used; tests which would have chosen a different path to use to step through the analysis. The test for a "speed" horse was one such test. Had one of the two contenders been a speed horse, our analysis would have then checked to see if there was another speed horse in the field. If so, we would then check our "speed" contender to see if he had ever won a race after being engaged in a speed duel, or if it had ever lost a race after being engaged in such a duel. If the latter had been true, we could have eliminated our speed horse contender in this test. As neither of the horses were speed horses, this test was never used.

Now that I have given you, hopefully, a feeling for how an "expert system" would step through and handicap a race, let me briefly discuss a more general, better, model which could be used in an expert system. This is the model which I briefly mentioned at the start of this section.

We can break the entire problem of handicapping down to several major tests: 1) How fast can the horse run in top condition, 2) Will the horse be in top condition for this race, and 3) How will other exterior factors such as jockey skill or speed horses influence whether our horse runs as fast as possible? As mentioned previously, there are several tests which may be used to determine the answers for each of these major test categories,. Our expert system could select the most successful test for this situation, based on its knowledge of the situation, and factors such as class and distance.

Let's look at the issue of "how fast". The answer to this lies in speed ratings. A speed rating is a measure of how fast a horse ran a race, normalized to the track condition, and the particular characteristics of the track. See the primer in Chapter 1 for more information about speed ratings. But which speed rating should we use? Possible choices are: 1) the speed rating for the horse's last race, 2) the average of the

speed ratings for all of the horse's past races, or 3) the average of the speed ratings for the horse's last two good races (Quirin style). Or maybe we should use the speed ratings plus or minus some small value, such as the standard deviation of all of the horse's speed ratings or the standard deviation of the speed ratings of all horses in this class range. The choice of which method to use to calculate "how fast", i.e. the choice of the way to calculate the speed rating, could be a function of such things as the class of this race, and the distance. In this way, our system could use the situation of the race to decide which test to use.

Similarly, our expert system could choose a test for determining "condition" by checking out the existing situation. Has the horse raced within 14 days? If not, has it had a workout within 14 days? Who is the trainer, and does he have a high percentage of successes with horses returning to race after a layoff? Has the horse been showing signs of improvement in his last few races (indication of improving condition)? Does the horse usually hold his condition for a season of racing? (This is what Quirin tries to determine in his test "Average Earnings per Start" - a horse which holds his condition should have a higher result to this test than one that doesn't.) Once again, our expert system can choose one or more of these tests, based on what it knows about the general condition of horses in today's class which usually race at today's distance.

Finally, we need to predict how our horse will "logistically" run the race, i.e. where he should be at each position call of the race, and how the other horses in the race will interfere which him. *[15]*

If our horse is a speed horse, then we would like to know if there are other speed horses in the race. If there are, can our horse withstand the challenge of being contested for the early lead? Has he won under these circumstances before? Or, is the opposite true? Is our horse a late-breaker, who typically runs being the pack for most of the race, and then makes his move on the final turn before the homestretch? If so, does he have a jockey on him, today, with the necessary skill to pilot him around the tiring horses in

front, while saving ground near the inside rail on the turn, so that the horse has a chance at winning? In the later case, jockey skill is essential; in the former case (with the speed horse), jockey skill is not so essential. Once the speed horse is out in front, the jockey only has to keep it going. By attempting to predict where the horse under consideration will be at different stages of the race, the expert system can decide how important certain tests (such as jockey skill) should be in the analysis.

This has been a second, more general example of how an expert system could be used to analyze a race. In this case, we identified the essential "major" categories of tests which we would use to handicap the race, and then used the expert system to decide which actual test(s) to use.

Referring once again to my example using the horses in the 1st race on March 20, 1985 at Suffolk Downs, I would point out one last interesting factor in this race. This is a feature which was overlooked by our first primitive expert system, but which might have been picked up by our second system, if we had also programmed some "sanity" checks on the quality of the contestants, in addition to our three major categories of tests.

That factor to which I refer, is in the horse "Irish Popular". This horse just suffered two miserable defeats in its last two races, the last race a route, and the one before that a sprint. Both of these defeats took place in a **lower** class. So why has the trainer chosen to race the horse today in a higher class, for which the horse shows no aptitude? Well, I don't know, but maybe I could find out. This is more then just idle speculation - the horse actually came in third in this race, and paid handsomely as a "show" bet! WHY?

Well, one of our tests in our expert system should check for this type of anomaly, i.e. a bad-looking horse moving up in class. And there is only one place to check. The database. Has the trainer of Irish Popular ever pulled this trick with other horses? More importantly, if so, has he been successful? These are the things which a truly

complete handicapping program must look for, if it is to find bets which other people (or programs) overlook.

I have presented, in the preceding paragraphs, an example of how an "expert system" type of handicapping model could be used to analyze a race. There is one other model I would briefly mention.

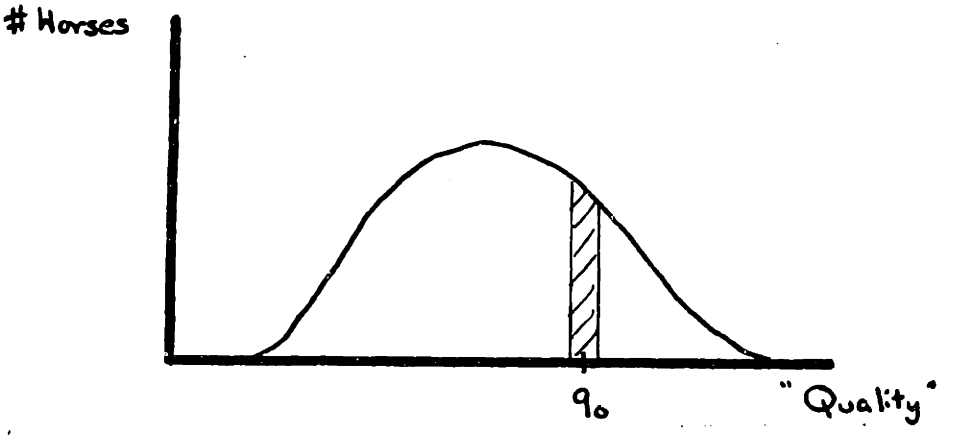
Suppose that I had a full database of horses on which to run tests and analysis. What would I check for? Well, trainer patterns and styles, such as that of the trainer of Irish Popular would certainly be one thing. But there are other analyses which could be carried out on a data base which would be very interesting.

Suppose that we took every horse in the database, and applied some test for the overall quality. This could be a combination of earnings, consistency, speed ratings, speed points, and other factors. We would have a distribution of horses of varying "qualities". Now suppose we wanted to check for things like the variability of speed ratings for a certain quality of horse. (The results of this type of test could have been used, for example in comparing Broadway Director with a speed rating of 98 with Go to Glory, whose speed rating was 97.) In our primitive expert system, we decided that horses with speed ratings within three points of the highest speed rating were fair game. But is this really true?

For each "quality" of horse, we could plot the average speed rating, and the variance. From such calculations, we could decide if three points was reasonable for a ballpark figure, or if four points or two points would be better. See Figure 14 for quick example of how this could be done. These variances would undoubtedly be different for horses of different quality. As mentioned previously, horses of higher class hold their conditioning longer, and hence should have more consistent speed figures. Care would have to be taken in compiling the data for this calculations, for it is often the case that when a horse is losing a race, the jockey will not push it. When this happens, the horse's speed rating for that race is not a true indicator of the horse's ability. The speed

Figure 14: Using Probability to Determine Range of Acceptable Values

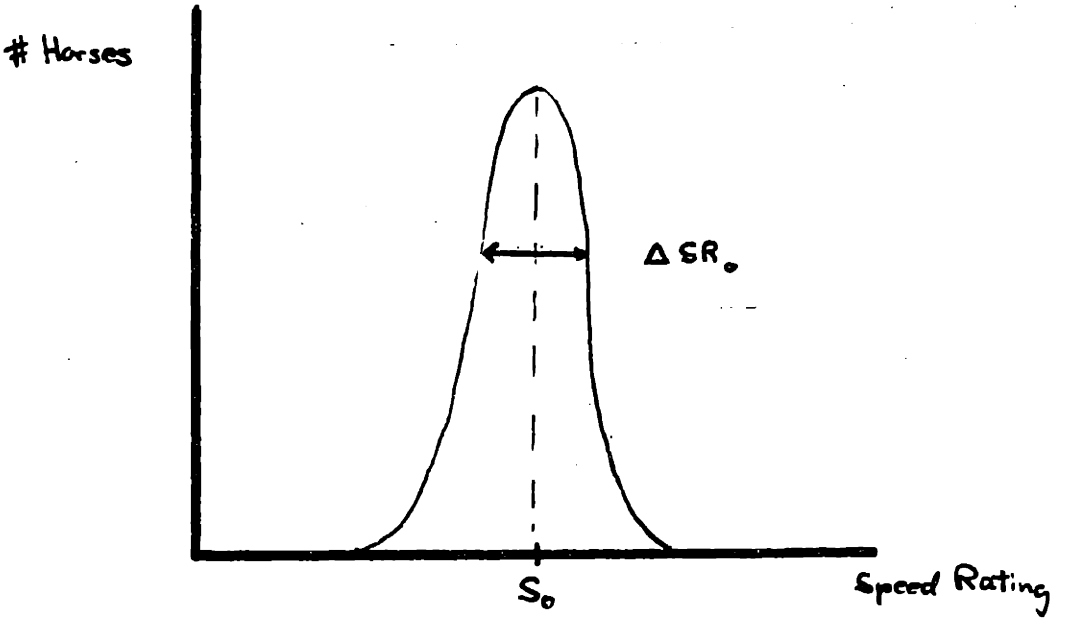
Ex. Let "Quality" = Point value returned by Quirins Sprint Formula



Pick a quality, q_0, of horse

For all horses of quality q_0, calculate some particular factor of interest. For this example, speed rating.

Then plot:



Determine S_0 and ΔSR_0 for q_0

Can use S_0, and ΔSR_0 to determine roughly the probable changes in speed rating of a horse of quality q_0.

ratings for these "very bad" races should not be included in our analysis of the variability of speed ratings. Subtleties such as this exist all through handicapping, which is why things like means and standard deviations must be at first taken with a grain of salt, unless they are made accurate by the careful consideration of which data points be used in their calculation. If such care is taken, however, they could prove to be powerful indicators of where the lines should be drawn in tests comprising an expert system.

This chapter has discussed the results of testing done on the WHINNY input and analysis systems. I have explored two alternate models to the linear formula approach, both using "expert" system ideas. In the more general of these models, tests representing different ways to identify the same concept are grouped into major categories. The expert system uses the conditions of the race and it's knowledge of the situation to select from these categories the tests which will be the most effective in this situation. In addition to just testing the qualities of the horse under consideration, our system must also try to predict what predicaments may happen to the horse during the race, and if the horse has the qualities necessary to overcome these predicaments. Finally, I have included some wishful thinking as to how an expert system could be used to handicap a race, and some discussion of how the study of a complete database could lead to the discovery of how to produce more accurate tests.

7. Future Possibilities for WHINNY

I have succeeded in developing a versatile, expandable, handicapping tool, known as the WHINNY system. This tool is not an end unto itself, however, but merely a beginning. I have, in this thesis, only begun to explore the many ways in which the WHINNY system can be used to facilitate thoroughbred handicapping, or to increase our understanding of how factors combine to help predict the outcome of thoroughbred races. In this final chapter, I would like to put forth some of the ways in which the WHINNY system could be improved, either for more successful betting endeavors, or for a more general widespread, large-scale analysis of handicapping angles.

Upgrade of Linear Handicapping Model

While I have combined my various handicapping angles using a linear regression formula, of weighted sums, this is certainly not the only way, and probably not the best way to combine or relate these separate pieces of data. At the current time, the WHINNY system will only accept linear combinations through it's input program. (This is all that the system prompts the user for, in the construction of the analysis objects.) This certainly is not the limit of what the system can handle. Because the results of each test are returned as a series of numbers, one for each horse, any possible mathematical combination of these numbers is possible. "Weightings" based on the square, or cube, etc. of these test values are possible, along with, say, such exotic combinations as the results of one test, divided by the results of another test. It would be easy to build this generality into the WHINNY system, and it might prove to be useful in the construction of additional models and in understanding of which forms make sense for which factors.

Implement New Handicapping Models

There are also many interesting analysis methods, to which I previously alluded back in Chapter 2, but never had a chance to implement or test. One such method, which I believe has a good chance of success, is an analysis of a database, with respect to trainer patterns (or owner patterns). Ideally, I would like to be able to instruct the WHINNY system to search through it's database, and collect all of the horses trained by a certain trainer. These horses then could be subdivided along lines, such as class, or usual racing distance (sprint or route). What would be interesting to look for would be patterns, such as: 1) number of losses between winning races, 2) relative odds for winning races, compared to odds for losing races, 3) jockey quality for those winning races, and/or changes in jockey before a winning race, or 4) number of workouts before a winning race or 5) the number of winners a trainer has on any horse's first race back after a layoff. (Some trainers are exceptionally good at conditioning a horse for an outstanding effort immediately after a layoff.) Thus, I believe that such an analysis of patterns, either among trainers or owners, would be a valuable predictive tool for handicapping thoroughbred races.

Another useful addition to the WHINNY system would be the continued programming of the rules set forth by Tom Ainslie. Although I had originally planned to program all seventy-seven of Ainslie's rules, I was unable to do this, due to time considerations. Therefore,one of the logical steps to improving the WHINNY analysis system would be the programming of these additional rules.

Of course, the best improvement to the WHINNY analysis system would be the implementation of some of the additional models which were discussed in Chapter 6. Either the "primitive" expert system, or the more general expert system approach would greatly enhance the effectiveness of the WHINNY analysis. Both of these systems would actually handle the non-linearities and interrelationships which exist both among handicapping angles and the physical horses running the race, instead of

just glossing over them. While the primitive expert system would be initially easier to program, the more general "expert" system allows the user more flexibility for either adding new tests to the analysis, or fine-tuning the existing analysis. One or both of these models would greatly influence the effectiveness of the WHINNY system for accuracy in the selection of the "best" horse, while providing the opportunity for better understanding of the interrelationships which exist in the handicapping situation overall.

Development of Betting Strategies

The idea of finding betting schemes was also one which, although discussed previously, was not actually implemented. The final results returned by the WHINNY analysis system are a set of numbers, one for each horse, which describe how well each horse scored on each test. Theoretically, the horse with the highest (or sometimes lowest, as in the Quirin systems) should be the winning horse. Yet some considerations should be made for a horse whose final points are either significantly better than the rest of the pack, or just slightly above average. Surely, the first horse, with the greater point difference between it and the pack would be a better bet than the horse which is just slightly above the rest of the pack! How should the user (handicapper) bet on either such horse? There must be some relationship as to how much money should be confidently bet on a horse, based on it's point difference above the rest of the field. If there are two outstanding horses, would it be better to bet both for win, both for place, one for win and the other for place, or bet a perfecta (first and second horse, in order) on the two horses, in either order? These are some of questions that most plague a handicapper, as he tries to determine how to bet, (i.e. in which betting pool(s) to put his money, and how much for each one.) The problem is also complicated in that the final odds for each horse for each pool are not actually known until post time. This changes the problem from a more straight-forward probability analysis into a strategy which

must allow updates as the odds change. The WHINNY system could be used, however, in an attempt to develop a betting system in which such updates are possible.

The basic way in which to develop a "betting strategy" in the WHINNY system is to use the system to distinguish horses with very similar characteristics in similar settings, and then to study how often horses in any one of these categories succeed (where a success may be either a win, a place, or a show, depending on your interest.) The success of such an analysis hinges on how much information is available in the database for use in a betting analysis. Given a large enough database, that which we are interested in is 1) some determination of the relative quality of horses for this race (i.e. the point value results of some WHINNY analysis), and 2) the odds (as determined by the bettors at the track that day) for each of the horses. For any such analysis of possible betting strategies, we will need a sizable WHINNY database, from which to glean these pieces of information.

Lets assume that we have such a database. One way to determine a betting strategy is to apply our knowledge of probability and decision analysis to the data we have stored. We would divide our pool of all horses first by their "quality", as determined by some WHINNY analysis, secondly, by their effective "quality ranking" in the field of horses for some race and, finally, by the point values surrounding the ranking. (These three divisions of data will greatly reduce the number of horses we have available for any particular probabilistic study; however, without these divisions, we would be mixing apples and oranges, by not isolating the qualities which we wish to base our betting strategies on. These qualities are, once again: 1) Quality, 2) Quality ranking, and 3) Difference in point values (between our horse, the horse ranked higher, and the horse ranked lower.))

Once we have a category of horses with these similar quantities, then it would be possible to determine how often this set of horses wins, places, or shows, i.e. the probabilities of these events happening. Assuming we could find these probabilities

(and means, and standard deviations) for each quality. ranking, and point difference, we could then combine this knowledge with the available odds for each of the betting pools (win, place, and show) to determine where the best spot would have been to "invest" our money, and how much. The results of such a study would be, that, given any horse with a certain "quality", in a certain ranking, with certain point differences between him and the next better and next worse horse, we could then determine at what odds we would consider betting on the horses. Taking this one step further, we could also use our information of the situation to try to determine the optimum combination of bets on a race to maximize our expected profit.

Use of Body Language

There is one additional aspect which can greatly increase the handicapping success of the WHINNY system, and the betting success of the aspiring horseplayer. It is not something, however, which can be programmed into a computer, and then used in the analysis of a race. Despite all of the formulation, and calculation, and statistical analysis, it still remains a fact that horses are animals, and not machines, with good days and bad days, some of which are not entirely their fault. It is possible, however, to take advantage of these good and bad days by being able to recognize their existence, and the ability to change your analysis accordingly. The way to recognize these conditions is to try to personally analyze the body language of the horses at the track, and to decide which horses are feeling good, and which ones are not. The key to learning to do this type of analysis is in *The Body Language of Horses*, a 1980 publication by Tom Ainslie (of course) and horse trainer Bonnie Ledbetter [Ains80]. For your convenience I have included as an appendix, a quick summary of the major points stressed in this book. A few minutes of study will enable anyone to understand what body language to look for in thoroughbred horse at the race track. The proper interpretation of body language is an invaluable skill, both for understanding why things happen in a race (when the

computer says they shouldn't) and for successful betting and the avoidance of costly errors.

So far, in this chapter I have presented you with some of the ways in which the WHINNY system could be improved for future use. I have also mentioned some of the ways in which this handicapping tool could be used to explore various aspects, such as trainer patterns or betting patterns, that would be tedious, if not impossible, to do by hand. By such improvements, one may improve the accuracy of the handicapping done by the WHINNY analysis system. In the end, however, personal judgment, and the effective use of body language are important tools to use when betting seriously at the track. Next, I would briefly discuss the WHINNY system, itself, and how I feel it fits into the handicapping evolution of the past three decades.

Summary of WHINNY

In this thesis, I have described the handicapping tool which I have built, called the WHINNY system. In Chapter 1, I presented an overview of what I intended to cover, and a short primer about thoroughbred racing. A review of the past efforts made in the evolution of serious, methodical handicapping was covered in Chapter 2, while Chapter 3 contained the methodology of my own stab at the problem, and described what the WHINNY system was, and how it operated. Chapter 4 describe what rating schemes I used to compare horses; these same rating schemes were combined into actual handicapping analyses in Chapter 5. Then, in Chapter 6, I presented the results of testing the WHINNY system, with separate testing done on both the input and analysis aspects of the system. Finally, Chapter 7 contained a short description of how the WHINNY system could be used or improved in the future.

The development of the WHINNY system took over five months of part-time programming effort, three-fourths of which were devoted entirely to the input system. I chose to implement the WHINNY system on the *Symbolics 3600* Lisp Machine, because of the features this machine had to offer, the most important being: 1) a window/menu system which allowed input to be accomplished via a pointing device on the screen, 2) the use of flavors which allowed both generalized message passing between objects of dissimilar types, and a hierarchical structuring of objects along with their inherent abilities, and 3) the interactive nature of the machine, which allowed new analysis to be created and implemented quickly and easily.

It is my opinion that the WHINNY analysis system, as it currently stands, only uncovers the "tip of the iceberg" of all that may be understood about thoroughbred handicapping. Dr. Quirin's formulae tend to point toward the obvious, in their handicapping of a race, and very often choose the favorite or the second favorite as the "best" horse. I feel this will prove to yield, after further testing, around 40% winners. (This is about the success rate of professional track handicappers, although some claim

to have hit 50% at one time or another.) However, the WHINNY system is capable of digging much deeper into the problem, especially with the implementation of new models suggested earlier in this chapter. WHINNY is built in such a manner that new models, even non-linear ones, can be easily implemented and tested, via the WHINNY input and analysis systems. If this feature of the WHINNY system is used to aid the development of new handicapping approaches and models, I believe that a success rate higher than 50% is possible in the future.

The WHINNY handicapping system was not designed to be an end unto itself, but rather a computer tool useful for the continued exploration and analysis of thoroughbred racing. I have designed and implemented the system to be as general as possible, while allowing analysis of the most minute details. Through it's use, I hope that handicappers, statisticians, and all persons interested in horseracing may achieve a better understanding of the individual factors present in thoroughbred handicapping, and how these factors may be successfully combined to predict the results of horseraces.

Appendix I - Body Language

This appendix consists of a short summary of the types of body language that may be witnessed in thoroughbreds at the race track. The information is a summary of some of the facts presented in Chapter 7 of *Body Language of Horses.* [Ains80] For additional information about body language in horses, especially thoroughbreds, you should consult this excellant publication.

The following categories of body language go from very good to bad. They enclude these types of characteristics: sharp, ready, dull, frightened, angry, overheated, cold, and hurting. There are also some descriptions of the body language of a horse which seems to like a muddy or sloppy track. Finally, there are the signs which indicate a drugged horse.

The Sharp Horse

All eagerness. Prances on toes, head tucked down toward chest, next arched, ears pricked forward. Tail may be held slightly upward.

May sweat, especially between rear legs. This is caused by excitement.

Looks physically healthy. Coat, mane, and tail gleam.

Restless. Sniffs air. May dance and wheel. Muscles may quiverin eagerness, as the time of race draws near (sign of panic in other horses that do not look proud, i.e. neck arched.)

In post parade, starts gallop easily. Usually, almost lunges into the canter, tail up, hind legs digging in, emphatically. May rear slightly. Sometimes drags the lead pony along, causing it's ears to go back in frustration. Jockey may stand up in saddle to control horse (or to make crowd believe that the horse is really sharp when it is not. Be wary)

The Ready Horse

Healthy and content. Quieter, fully tractable, and less hectic. Walks out willingly.

Probably will not sweat at all, before the race.

Coat may gleam, but does not match that of a sharp horse.

Moves nicely, neck not arched. Jockey and handler are also much more relaxed.

In post parade, lead pony probably will canter before the thoroughbred does. Then thoroughbred moves smoothly into a slow, collected canter with head down. Jockey will not have to stand up to control.

The Dull Horse

Moves willingly, but in a flat-footed walk without any spring. Neck not arched. Not interested in crowd. Relaxed. Ears sometimes flopping sideways, like airplane wings.

Sweat along neck, but not between rear legs.

Coat seldom gleams, and may even look rough.

In post parade, may walk flat-footed, or even shuffle. Jockey may try to wake horse by rattling bit, or shifting in saddle. When lead pony canter, the race horse only musters an awkward trot. Canters when tugged along. Head down, but neck not arched.

These are the horses that win nine out of every ten races. A ready horse may beat a sharp one, or a dull one may beat a ready one, but only if there is a substantial difference in their qualities, as discovered by handicapping. Most races at Suffolk Downs are won by ready or dull horses, as a truly sharp horse is hard to find. Now for the other types of horses one may find on a race track.

The Frightened Horse

A horse may be frightened at any time during the pre-race formalities. A sharp, ready or dull horse may quickly become frightened before the race; if it is not calmed or soothed by its handlers quickly, is has become a "frightened horse".

Sweat at neck, shoulders, and between both pairs of legs.

Head high, moves rapidly, eyes rolling violently, nostrils flaring, ears constantly flicking. May whinny or neigh for help. Leg action high and erratic, tail swishs from side to side, or up and down, but does not "pop", like an angry horses.

Handler has a tight hold on horse, sometimes with a chain over the nose. Horse moves in a half-circle, fighting the chain, and trying to flee.

Reacts strongly to any sudden noise or movement. Shies from the crowd. Will probably enter saddling enclosure eagerly, seeking sanctuary. Sweat increases and teeth may chatter. Often seen led out of the stall, circled in the paddock, and led back in, in an attempt to use up some of the restlessness. Extra people come into stall to "help".

In post parade, jockey has a tight hold. Horse may calm when next to lead pony, but too much energy has been lost. Tries to touch as much of lead pony as possible, and often holds its head over the lead pony's neck. Nose is high, ears flicking, legs high and erratic, hind legs well underneath, moves at small spastic jumps. May fight at the starting gate.

The Angry Horse

These descriptions cover anything from mild irritation to wild fury. If only irritation, may still have a chance at winning, if the superior animal. Otherwise, probably not.

Rerely sweats.

Ears flatten back against the head at times of annoyance, else straight forward. The longer time spent flattened, the worse the anger. Eyes often in a fixed stare, at object of anger. Upper lip may curl. Legs move delibrately, tosses head in attempt to break lead rope.

May pop tail, i.e. raise it, and sharply bring it down against hind quarters. Or may rapidly swish it from side to side, like a whip.

If saddling is the annoyance, it will cause a fight in the saddling area. Or if the jockey is the problem, the angry will appear when the jockey does. Goes to walking ring ina tense quick trot, leaving very deep hoofprints. May try to bite the jockey or the handler.

In post parade, may try to bite handler or lead pony. Sometimes angry horses are excused from the post parade, and start an early warmup, to burn off some of their anger.

The Overheated Horse

Moves as slowly as possible. Lethargic. Sweat all over.

May kick a hind legs, as if in anger. Really to shake off dripping sweat.

The Cold Horse

Teeth chatter, eyes half closed, ears flopped, and head is down. May even get annoyed when jockey mounts.

The Hurting Horse

Coat is very dull. Eyes are dull. Motions are subdued. Moves in a slow, collected manner. Head hangs and ears droop. Tail hardly moves. Feet shuffle.

In post-parade, the lead pony always leads. Horse reluctantly changes into a faster gait. In trot, the head droops very low. In canter, the head may bob along, in an attempt to change balance off of hurt member. Canter is awkward, stilted and choppy.

Horse may warm out of discomfort, much as athletes do. This will become evident by a raised head, and a smooth stride.

The Mud Horse

Any horse behaving contently in heavy rain. Not slouching around with eyes half shut, and ears down and back to keep out the rain.

In post-parade, it moves right out, instead of hiding next to the lead pony. Plants feet firmly and naturally in the mud. Moves deliberately. Uncomfortable horses act as though walking on eggshells or marbles. They pick up feet quickly, dance sideways, and maybe have ears back. Mud horse moves with balanced, determined stride.

As a mentioned previously, there still are a few cases where a trainer or owner will drug a horse, either with amphetamines or barbituates, to influence its perfomance. As with drugged humans, there are certain behaviors to look out for.

The Down Drugged Horse

Ears do not function properly. They flop sideways in the airplane position, open toward the ground. When the animal reacts to something, the ears to not follow, or point at the object, as they would in a normal horse.

Under a less-severe dose, the ears may move in coordination with the rest of the body, but sluggishly. They often return to the airplane position, and may actually flop up and down.

Head bobs low, feet shuffle, and tail moves slowly. May drool, or walk into another horse.

The Up Horse

Ears do not function properly. They freeze in whatever position they were in when the drug took effect. Usually they prick forward, and remain that way.

All motions may become hyperactive. Unusual head movements, erratic dancing, and muscular twitching. Legs move unusally high, and come down stomping at the ground. Breathing is rapid.

Eyes may appear to be "spaced" out. May foam at the mouth. Base of the tail may stick out.

Under a less-severe dose, the ears may move somewhat. They always return to the original position, however.

The handicapper may someday be faced with the choice of trying to decide whether or not to bet on a horse that looks like it has been stimulated. While the trainer or the owner are betting that the horse will perform better than usual, this may only happen if it doesn't hurt itself before or during the race. My own reaction, in such a betting situation, is to pass the race.

There is a drug, called "Lasix", which is legal in Massachusetts, Maryland, and some other states, not including New York. This drug is similar to one used by many human long-distance runners, and it controls bleeding in the lungs, which hampers a horse's breathing. The fact the a horse is on Lasix is indicated in the racing program by the letter "L" after the horse's name. Also, if the horse is on Lasix for the first time, this will also be indicated, at the bottom of the program. A horse on Lasix for the first time

often performs much better than in previous races. When a horse use to Lasix competes in New York, or another state which does not allow the drug, it's performance will inveriably suffer.

Endnotes

[1] A wide range of such handicapping wonders can be found advertised in the daily periodical, the *Daily Racing Form*, published by *The Daily Racing Form, Inc.* and available at many local newsstands.

[2] For a more complete description of everything about thoroughbred racing, an excellent reference is the appendix to *Ainslie's Complete Guide to Thoroughbred Racing*, 1968, listed in my references.

[3] In the past, as a result of English racing heritage, fourth was also considers "in the money", but in most American races today, fourth is not considered "in the money"

[4] The Triple Crown for Thoroughbred Racing is a series of three races for outstanding 3-year-olds. The three races are: The Kentucky Derby, The Preakness Stakes, and The Belmont Stakes.

[5] I have programmed WHINNY so that it doesn't matter which mouse buttons is clicked, to secure the data.

[6] We tried these on menus, and found that in this case, typing on the keyboard was faster.

[7] This analyses are valid subcomponents of any analysis, as any analysis can have as many layers as is necessary. More on this in Chapter 5.

[8] Dr. Quirin's formulae are constructed as to give the lowest score to the best horse.

[9] This is most likely not a very linear quantity, and probably would do better if it was treated in a different manner (or model).

[10] There may be more than 8 past races entered for a horse, and it is possible that the horse could have failed in more than eight races.

[11] As may be expected, this advantage also depends on how many speed horses there are in the field. Two or more speed horses often burn themselves out fighting for the lead, leaving the winning to a horse which was willing to hang behind for most of the race, and then overtake the tiring leaders. This is definitely not a linear effect.

[12] However, at Saratoga race track, on a very muddy day last summer, I witnessed a young three-year-old colt break the existing track record! Mud does not necessarily stop some horses.

[13] While the weight of 250 points seems rather high, it is very often the case that the jockey is the deciding factor in a race (due to racing "luck", and the skill of the jockey). In other cases, however, there is a clear cut "best" horse, which could probably win despite a poor jockey. This does not seem like a linear effect. However, it is also the case that trainers often change to a better jockey for a race which they fully intend to win. In this case, the "top-5" jockey is the best indication of the horse's current condition, and the stable's intentions. So maybe 250 points is not really excessive.

[14] This is also a good place for a non-linearity. Perhaps in a non-linear model, this could be checked for, and predicted.

[15] Trainers often plan these details right down to the length, and often create very interesting strategies for running the race. In some races, "entries" are permitted - these are two horses running physically in the race under one number. If either horse wins, a bet on that number "wins". The strategies involved with entries are often incredibly complex. For example, in one case, one horse is suppose to break fast, set a deadly pace, and then fade back to the outside, allowing the second horse of the entry to come up along the rail, and win the race. Talk about teamwork!

Bibliography

[Ains68] Ainslie, Tom, *Ainslie's Complete Guide to Thoroughbred Racing*, New York, Simon and Schuster, 1968.

[Beyr75] Beyer, Andrew, *Picking Winners: A Horseplayers Guide*, Boston, Houghton Mifflin Company, 1975.

[Davd77] Davidowitz, Steven, *Betting Thoroughbreds: A Professional's Guide for the Horseplayer*, New York, E. P. Dutton, Inc., 1977.

[Beyr78] Beyer, Andrew, *My $50,000 Year at the Races*, New York, Harcourt Brace Jovanovich, 1978.

[Qurn79] Quirin, William, Ph.D., *Winning at the Races: Computer Discoveries in Thoroughbred Handicapping*, New York, William Morrow and Company, Inc., 1979.

[Ains80] Ainslie, Tom, and Bonnie Ledbetter, *The Body Language of Horses*, New York, William Morrow and Company, Inc., 1980.

[Beyr83] Beyer, Andrew, *The Winning Horseplayer: A Revolutionary Approach to Thoroughbred Handicapping and Betting*, Boston, Houghton Mifflin Company, 1983.