

HEURISTIC CLASSIFICATION FOR HUMANS WITH PREFRONTAL CORTEX DAMAGE: TOWARDS AN EMPIRICAL MODEL OF PHINEAS GAGE

By

Daniel Houser Kevin McCabe

Interdisciplinary Center for Economic Science
George Mason University

Michael Keane

Department of Economics
Yale University

Antoine Bechara

Department of Neurology
University of Iowa

April, 2003

Abstract: In many research contexts it is necessary to group experimental subjects into behavioral “types.” Usually, this is done by pre-specifying a set of candidate decision-making heuristics and then assigning each subject to the heuristic that best describes his/her behavior. Such approaches might not perform well when used to explain the behavior of subjects with prefrontal cortex damage. The reason is that introspection is typically used to generate the candidate heuristic set, but this procedure is likely to fail when applied to the decision-making strategies of subjects with brain damage. We suggest that the Houser, Keane and McCabe (HKM) (2002) robust behavioral classification algorithm can be a useful tool in these cases. An important advantage of this classification approach is that it does not require one to specify either the nature or number of subjects’ heuristics in advance. Rather, both the number and nature of the heuristics are discerned directly from the data. To illustrate the HKM approach, we draw inferences about heuristics used by subjects in the well-known gambling experiment (Bechara, Damasio, Damasio and Anderson, 1994).

Introduction

The unhappy circumstances of Phineas Gage are by now well known. Briefly, as related by Antonio Damasio in his *Descartes' Error* (1995), Gage was working as the foreman for a railroad construction team in Vermont in 1848, when an explosion blew an iron bar through his left cheek, skull and the front of his brain. The bar exited the top of his head at high speed, and Gage managed to survive the blast. Although his post-accident IQ remained high according to standard measures, he nevertheless underwent radical personality changes and, perhaps more interestingly, seemed to lose the ability to make good decisions. In particular, he systematically made decisions that were, by any objective measure, not in his long-run best interest. He eventually lost his job and family, and spent much of the rest of his life working as a sideshow attraction for a circus.

We now know that Gage suffered damage to the ventromedial (VM) area of his prefrontal cortex (see, e.g., Damasio, Grabowski, Frank, Glalburda and Damasio, 1994). People with damage in this area typically maintain good memory and score well across a wide range of personality and intelligence tests. However, they tend to have difficulty in making “good” decisions. That is, they often make decisions that seem clearly contrary to their best interest, even when they claim that they know this is the case.

Investigating the natures of the differences between VM and normal decision making has proved challenging, because VM patients perform as well as normal patients on many standard diagnostic tests. However, Bechara et. al. (1994) describe one laboratory experiment in which VM patients perform remarkably differently than control subjects. This experiment has been dubbed the “gambling task,” because it involves turning over cards sequentially and earning and losing money, according to the markings

on each card. Bechara et. al. (1994) report that VM patients choose cards from “bad decks” systematically more often than people without such brain damage. In their experiment, a bad deck is one that yields high immediate rewards but higher future losses, so that on average a person playing a bad deck will lose money. A good deck, on the other hand, provides lower immediate rewards but even lower future costs, so that on average a person drawing from the good deck will earn money. The main result reported by Bechara et. al. (1994) is that about 60% of VM patients draws are from bad decks, while this is true for only about one-third of their control subjects.

Bechara, Tranel and Damasio (2000) investigate three reasons, not mutually exclusive, for differences in behavior in the gambling task. These are that VM patients might be relatively (i) hypersensitive to reward; (ii) insensitive to punishment; or (iii) insensitive to future consequences. To discriminate these hypotheses they designed a new experiment, a variant of the gambling task, such that the bad decks yield low immediate punishment and even lower future earnings, while the good decks yield high immediate punishment and even higher future reward. Analysis of this experiment’s data allows them to conclude that neither (i) nor (ii) is supported by the experimental data, and that (iii) is a simple hypothesis consistent with the evidence.

In this paper we provide an alternative procedure for drawing inferences about the heuristics used by VM and control patients when playing the original gambling task. Our approach is to analyze data from the original environment using the statistical classification algorithm suggested by Houser, Keane and McCabe (2002). The goal of our analysis is not to provide new results about the behavior of people with VM damage. Indeed, experimentation over the last decade by Bechara and others has expanded the

knowledge of VM behavior far beyond what one can expect to gain by a statistical analysis of a relatively old data set. Rather, in this paper we demonstrate that the Houser, Keane and McCabe (HKM) classification procedure can be used to discern behavioral patterns that were not originally teased out of this data set, and that those patterns line-up well with what subsequent experimentation has already discovered.

A important goal of this paper, therefore, is to use the well-known gambling task data as a vehicle to highlight the potential advantages of an HKM analysis. There are several reasons that behavioral researchers in all fields, including economics, psychology and neuroscience, should be interested in the HKM statistical approach. One is that HKM does not require the researcher to pre-specify the nature or number of the heuristics used by subjects. This is in marked contrast to many approaches to type-classification that require the investigator to pre-specify the universe of possible decision rules (e.g., the popular strategy suggested by El-Gamal and Grether, 1995). Especially when analyzing the behavior of people with brain damage, it seems likely that the usual introspective process that generates this universe may fare quite badly.¹ In addition, HKM does not require that all subjects with a particular brain condition (in the present case, VM and control subjects) use the same heuristic. As discussed below, the idea behind the procedure is to group subjects according to similarities in their decision-making behavior, regardless of any known physical abnormalities they might possess.²

The data set analyzed in this paper is relatively small and unbalanced. It consists of 17 VM patients, and eight lesion control subjects who have brain damage in an area

¹ While very common, introspection is certainly not the only procedure available to determine a universe of possible heuristics. For example, objective evidence from neuroeconomic studies of decision making might provide useful insights into the cognitive strategies used by both brain damaged and normal subjects. See, e.g., McCabe, Houser, Truand, Ryan and Smith (2001).

² Of course, the researcher may incorporate this information into the HKM classification procedure.

outside of the ventromedial prefrontal cortex (in particular, to the left-somatosensory cortex.) Nevertheless, we show that an analysis can be conducted that groups subjects according to similarities in their heuristics, and that allows inference with respect to whether these heuristics differ in terms of their sensitivity to punishment (losses).

We allow for two types of heuristics in our population. Our results indicate that 15 VM patients and two controls use one type of heuristic, while two VM patients and six controls use the other. The two heuristics do not differ with regard to the way they respond to losses, which lines up well with the results of subsequent experimentation reported by Bechara et. al. (2000).

2. Statistical Methodology

The statistical procedure used in this paper is developed in detail in Houser, Keane and McCabe (2002), and will not be repeated here. Papers that discuss closely related procedures for inference in multinomial choice frameworks include Geweke and Keane (1999a), Geweke, Houser and Keane (2001) and Houser (2003). The HKM approach is useful whenever an investigator is interested in drawing inferences about the nature of behavioral heterogeneity in a population, but does not want to take a strong stand with respect to the nature of that heterogeneity. In particular, under relatively weak assumptions, the HKM algorithm draws inferences about both the nature and number of heuristics (or, equivalently, decision rules) used by subjects in a given population.

Loosely speaking, a decision rule is a map from information to action. For example, if people sitting in a theatre are given the information that the theatre is burning, many will likely decide to act by leaving the building. Behavioral heterogeneity might

exist even here: a few might decide to stay. Intuitively, the HKM approach allows one to draw inferences about both the nature and number of relationships that exist between the information people have and the actions they take, at least within a given context.

While many interesting types of decisions are easily observed, it is usually the case that the information that resulted in a particular action is not. This is less the case in laboratory experiments. There, much (often most) of the information that is relevant to subjects' laboratory decisions is under the control of, and therefore known to, the researcher. We exploit this control to specify the form of the heuristics that we investigate below.

3. The Gambling Task

Bechara's gambling task (Bechara et. al, 1994) is a sequence of static decision problems under ambiguity. The experimenter begins by giving a subject \$2,000 in play money. The experimenter places four decks of cards in front of the subject, and tells him/her that they can earn more play money by turning over cards, and that his/her goal is to earn as much play money as possible. The subject is told that every card they choose will result in them earning some amount of money, and that there will be occasional cards that impose costs on them. The subject is told nothing else. The subject then begins turning over cards, one-by-one, until they are told to stop by the experimenter. The stopping point is after 100 cards have been selected, although the subject does not know this in advance.

The subject is told nothing about the payoff or cost distributions within any of the decks of cards. In fact, the decks have been constructed in a very particular way. The

first two decks, call them A and B, provide a positive payment of \$100 for each card. However, they also have occasional very high costs. On average, turning over 10 cards in the A or B decks will have a net cost of \$250. The C and D decks have lower rewards per card, \$50, but also have lower occasional costs. On average, turning over 10 cards in the C or D decks yields a positive return of \$250. For this reason, we will refer to decks A and B as the “bad” decks, and C and D as the “good” decks.

The main result reported by Bechara et. al. (1994) is that VM patients choose from the bad decks statistically significantly more often than normal subjects. On average, around 60% of all VM patients’ draws are from the bad decks, while this is true of only about one-third of the normal patients’ draws. This led to much speculation about the source of the behavioral difference. One question was whether VM patients were relatively insensitive to losses, and if this insensitivity could explain the difference. Subsequent research by Bechara et. al. (2000), which used a new experiment designed to address this question, suggested that differences in loss aversion-behavior were not likely the source of the different choices. The results we report below provide convergent evidence for this conclusion.

4. The Model

The Houser, Keane and McCabe approach to type classification (2002) requires that subjects’ relevant information sets, and the link between information and action, be specified. We assume that each subject has a subjective “value” associated with draws from each deck of cards, and that they draw a card from the deck on which they place the highest value. The way that values are formed can be modeled in any way that the

researcher chooses, subject only to usual identification issues. In this chapter, because our intent is primarily illustrative, we use a very simple model that nevertheless captures important features of subjects' actual play. Denote the deck by j (with total number of decks J), the subject by n and the current draw by t . Assume that subjects assign values to draws in H different ways (that is, there are H valuation heuristics used in the population.) With this notation, we model the subjective value that subject n assigns to drawing a card from deck j at round t , assuming they use heuristic h , as:

$$\begin{aligned}
 V_n(j,t;h) = & b_{1jh} \\
 & + b_{2jh}I(\text{Last Draw was from deck } j \ \& \ t > 50) * \text{Loss}(t-1) \\
 & + b_{3jh}I(\text{Last draw was from deck } j \ \& \ t > 50) * \text{Reward}(t-1) \\
 & + e_n(j,t;h),
 \end{aligned}$$

where e is an identically and independently distributed Gaussian random variable that represents idiosyncratic noise, which might be related to failures to implement the heuristic perfectly. Because this is a situation of ambiguity, the model assumes that the subject uses the first 50 draws to gain experience in each deck. Inferences with respect to loss and reward effects are based on the final 50 draws experienced by each subject. Finally, the function "I()" represents an indicator function that takes value one if the condition inside the brackets is true, and is otherwise zero. This model simply posits that the value a subject places on drawing from deck j depends on a constant, noise, and his/her most immediate previous experience with that deck.

It is possible to use the HKM algorithm to draw inferences about the number H of heuristics in the population, the nature of each heuristic h in H (that is, the coefficient values), and to determine the probability with which each subject uses each heuristic. A specific way to do this is detailed in Houser, Keane and McCabe (2002), and involves a Bayesian analysis of a mixture of probits model (for more on mixtures of probits see, e.g., Geweke and Keane, 1999b).

For this paper's purposes, however, we take a dogmatic stand that there are exactly two heuristics at use in the population. There are two reasons for this decision. First, the results of substantial previous research with this population suggest that there are in fact two types of behavioral heuristics in this population, and using these previous results to inform our current model is reasonable. At the same time, note that there is no necessary reason to expect that all VM patients will follow the same heuristic, or that all normal controls will follow the same heuristic. For example, some VM patients might follow a strategy that looks very similar to the control subjects. The HKM procedure allows for this and other possibilities.

A second reason to assume that there are two types of decision rules in this population is that, as a practical matter, it would be difficult to interpret the finding that there are three or more heuristics in the population. The reason is that our sample size is rather small (8 controls and 17 VM patients), and evidence of more than two heuristics might not be robust to a larger sample, or the nature of the heuristics that we estimate might be a quite biased reflection of the true heuristics at use in the population, given the relatively small number of subjects that would be assigned to each.

This highlights an important feature of the HKM approach to type classification. Because it is a robust approach, in the sense that both the nature and number of heuristics are determined endogenously, it can be less efficient than procedures that take a stand on the heuristics that subjects use. Of course, if such a stand is wrong, and the model consequently misspecified, then the efficiency gain will come at the cost of specification error bias.

4.b. Implementation and Identification

Although there are two “good” decks, and two “bad” decks in the actual experiment, in this chapter we report results based on a model that treats each pair as one. Equivalently, we model the individual as making a choice between choosing a deck with \$100 payoffs or \$50 payoffs, and then randomizing across the two decks within that choice. Hence, we set $J=2$, which turns out to mean that there are three identified coefficients, along with one variance term with a pegged value, that characterize each heuristic.

To see this, note that the value function described above requires both location and scale normalization for identification. Location normalization is achieved by differencing:

$$\begin{aligned}
 V_n(1,t;h) - V_n(2,t;h) = & b_{11h} - b_{12h} \\
 & + b_{21h}I(\text{Last Draw was from deck 1 \& } t>50)*\text{Loss}(t-1) \\
 & - b_{22h}I(\text{Last Draw was from deck 2 \& } t>50)*\text{Loss}(t-1) \\
 & + b_{31h}I(\text{Last draw was from deck 1 \& } t>50)*\text{Reward}(t-1) \\
 & - b_{32h}I(\text{Last draw was from deck 2 \& } t>50)*\text{Reward}(t-1) \\
 & + e_n(1,t;h) - e_n(2,t;h).
 \end{aligned}$$

Note that the differenced constants are not separately identified, but are estimated as a single constant. Similarly, the differenced error component is treated as a single noise term. Also, because the nature of the experiment induces little variation in rewards, the coefficients on lagged rewards are only weakly identified. Consequently, we choose to drop them for the remainder of our analysis. Finally, scale normalization is achieved by pegging the variance of the error at a fixed value.

As a technical note, it turns out that to implement the Bayesian version of HKM as described in Houser, Keane and McCabe (2002) one must specify priors on the coefficients “b” that appear in the value expression above, along with the fraction of each type that exists in the population. We follow Houser, Keane and McCabe (2002) and use Gaussian priors with means of zero and standard deviations of one for the intercepts, and 0.1 for the coefficients on losses. We use a diffuse Dirichlet prior centered at $\frac{1}{2}$ for the fraction of each type in the population. These priors are quite weak relative to the posterior distribution. That is, the data is responsible for the results we report below.

5. Data and Results

Our data set consists of 25 subjects who played the gambling task one time. 17 of our subjects are VM patients, and 8 are lesion controls with damage to the left somatosensory cortex. The data were collected by Antoine Bechara and colleagues at the University of Iowa, and represent a subset of data that has been previously published in various books and journals. Figure 1 compares the frequency with which the two types of patients drew from the “bad” decks (the \$100 decks.) As has been previously reported, VM patients draw from the bad decks statistically significantly more often than the normal patients.

Moreover, as seen in Figure 2, the rate at which VM patients draw from the bad deck seems roughly constant over the entire experiment. The rate at which LC's draw from the bad deck is similar to the VM rate over the first 10 or so draws, but then declines substantially, but stays roughly constant over the last 80 or so draws.

5.b. Results

Our results are derived through the use of a Gibbs sampling algorithm. The Gibbs sampler is a recently developed numerical procedure for drawing inferences about statistical models like ours. In the present case, and as a technical detail, our inferences are based on a Gibbs sampling algorithm that we coded in FORTRAN 77 and that makes extensive use of IMSL subroutines.³ We ran the sampler for a total of 500 cycles. The results reported below are based on the last 250 cycles.⁴ Specifics on the way in which a Gibbs sampler can be implemented to draw inferences in this environment are available in Houser, Keane and McCabe (2002).

Consider first the way in which we type-classify subjects. We based our subject classification on the posterior mean probability that they were each type. If the posterior probability of being the VM type is greater than or equal to 0.495,⁵ then they are classified as that type. Otherwise, they are classified as the lesion control type. Note that this classification procedure assigns people to the VM type unless there is “reasonable” evidence to the contrary. This reflects our prior knowledge that the sample is unbalanced

³ Our FORTRAN code is available on request.

⁴ Visual inspection of the draw sequences suggested that convergence had been achieved by cycle 250. Complete draw sequences are available from the authors on request.

⁵ Three subjects had a posterior probability of 0.496 of being the ventromedial type, and in each case we assigned them to the ventromedial type. Two of these three are actually VM, so reversing the classification for these three reduces the number of total VM types by 3 (from 17 to 14), and induces one additional type classification error (from 4 to 5).

in favor of VM patients. (Alternatively, this prior could be incorporated directly in the algorithm, but this alternative approach seemed more transparent for the purpose of this paper.) The posterior type probabilities favored one type over another by only a few percentage points for most of our subjects. The highest posterior mean probability across all subjects of being the VM type was about 66%, and the smallest was about 36%.

Table 1 provides the results of our typing procedure. We call one of the two estimated heuristics the “VM” heuristics, simply because most of the subjects assigned to it are VM patients. We denote the other heuristic as the “LC” heuristic for the same reason. It turns out that 17 subjects are classified as VM types, which is identical to their frequency in the data. However, two subjects classified as VM are in fact lesion controls. Thus, four subjects are “mislabeled,” in the sense that their actual brain condition is not reflected by the label of the heuristic that they use.

Table 2 describes the marginal posterior distributions for the coefficients of each heuristic. Notice first that the marginal posterior distributions of the coefficients for the amount lost in the previous period have the majority of their mass to one side of zero for both the LC and VM heuristics. Moreover, the values of these coefficients are very similar. This suggests immediately that, as reported by Bechara et. al. (2000) based on a different experimental design, both VM and lesion control patients respond to losses incurred in the previous period, and that these responses are similar. On the other hand, the posterior means of the constant terms for the two heuristics differs by about 0.1, and the posterior means lie on different sides of zero. This provides some evidence that the baseline rate at which VM’s and LC’s choose from the bad deck differs.

The standard deviation of each heuristic's error term is 0.96, and this can be used to provide an interpretation of the coefficient estimates.⁶ Evaluated at posterior means, subjects using the VM heuristic will choose from the bad deck at a baseline rate of about 52%, which is about 10 percent higher than the LC's 48% baseline rate. These baseline rate differences are compounded by the effect of losses. Our estimates imply that both types of subjects are more likely to choose from the same type of deck after experiencing a loss in that deck, than they would be otherwise.⁷ For example, a subject who turns over a card in the "bad" deck and receives a cost of \$1,250 will choose from the bad decks again with 82% probability if they are using the VM heuristic, and 77% under the LC heuristic. Experiencing a cost of \$250 from the "good" decks generates probabilities of 17% and 16%, respectively. Overall then, these results indicate that VM patients choose cards from the bad decks at a higher baseline rate than the LC subjects. As a result, they experience losses more frequently, and these losses lead them to choose yet more frequently from the bad decks. The interaction of higher baseline choice rates and the effect of experiencing losses lead to the substantially higher bad deck choice frequencies of VM subjects.

Our analysis leaves unanswered the question of why VM patients would tend to choose from the \$100 decks at a higher baseline rate than the lesion controls. Further experimentation by Bechara et. al. (2000) suggests that the reason may be that VM damage leaves one unable to assess the future negative consequences of one's actions

⁶ The value of this parameter varied slightly during the Gibbs sampling algorithm due to a technical oversight. This variation was quite small and exhibited no drift.

⁷ This might be counterintuitive. One possible explanation for this behavior is that subjects come to expect that it is unlikely to experience two losses in a row in a given deck. Alternatively, this result might reflect an aggregation effect embedded in our statistical model. In particular, it is possible that subjects are in fact switching decks after a loss, but not switching reward amounts.

accurately. Certainly, this reasoning is consistent with the life and times of Phineas Gage, and is supported by the skin-conductivity experiments reported in Bechara et. al., (1997). VM patients seem to react to negative events, and yet not act in the future as though they were aware that more such events might occur.

6. Concluding comments

Uncovering the nature and number of behavioral heuristics that people use, even in very narrow contexts, presents one of the most important current challenges to the behavioral sciences (see Houser, 2002b for an elaboration of this point.) A standard approach to this involves somehow determining a universe of possible ways that people might act, and then determining which one among this universe fits each person's behavior best (see, e.g., El-Gamal and Grether, 1995). While this approach has been shown to work well in many circumstances (see, e.g., Houser and Winter, 2003), there are some environments in which its success is less likely. The study of brain damaged people is one such environment, because it does not seem likely that introspection by a person with a normally functioning brain could provide accurate guidance on the heuristics that might be used by someone with a significant brain abnormality. The HKM classification procedure is a robust alternative, and we have demonstrated in this paper that it seems to work reasonably well within the context of the well-known gambling task environment. In particular, the results obtained by the HKM statistical procedure line-up well with the results of broader, subsequent experimentation with VM patients.

References

- Bechara, A., D. Tranel and H. Damasio. 2000. Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain*, 123. 2189-2202.
- Bechara, A., A. R. Damasio, H. Damasio and S. Anderson. 1994. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, 50. 7-12.
- Bechara, H. Damasio, D. Tranel, and A. R. Damasio. 1997. Deciding Advantageously Before Knowing the Advantageous Strategy. *Science*, February 28, 1997; 275(5304): 1293 - 1295.
- Damasio, A. 1994. *Descartes Error*. G. P. Putnam's Sons. New York.
- Damasio, H. T. Grabowski, R. Frank, A. M. Galaburda, and A. R. Damasio. 1994. The return of Phineas Gage: The skull of a famous patient yields clues about the brain. *Science*, 264. 1102-05.
- El-Gamal, M. and D. Grether. 1995. Are people Bayesian? Uncovering behavioral strategies. *Journal of the American Statistical Association*, 90, 1137–1145.
- Geweke, J. and M. Keane. 1999a. Bayesian inference for dynamic discrete choice models without the need for dynamic programming. In Mariano, Schuermann and Weeks, (eds), *Simulation Based Inference and Econometrics: Methods and Applications*. Cambridge University Press. (Also available as Federal Reserve Bank of Minneapolis working paper #564, January, 1996.)
- Geweke, J. and M. Keane. 1999b. Mixture of normals probit models. In Hsiao, Lahiri, Lee and Pesaran (eds.), *Analysis of Panels and Limited Dependent Variable Models*, Cambridge University Press, 49-78. (Also available as Federal Reserve Bank of Minneapolis staff report #237).
- Geweke, J., D. Houser and M. Keane. 2001. Simulation based inference for dynamic multinomial choice models. In B. Baltagi (ed.), *Companion to Theoretical Econometrics*. Blackwell. 466-493.
- Houser, D. 2003. "Classification of Types of Dynamic Decision Makers," in Nadel, L. (ed.), *Encyclopedia of Cognitive Science*, Vol. 1, Nature Publishing Group, London. 1020-1026.
- Houser, D. and J. Winter. 2003. How do behavioral assumptions affect structural inference? Evidence from a laboratory experiment. *Journal of Business and Economic Statistics*. Forthcoming.

Houser, D., M. Keane and K. McCabe. 2002. Behavior in a dynamic decision problem: An analysis of experimental evidence using a Bayesian type classification algorithm. Working paper, George Mason University.

McCabe, K., D. Houser, L. Ryan, V. Smith and T. Trouard. 2001. A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Science*. 98:20, September. 11832-11835.

Actual Brain Condition

	VM	LC	Total
Classification	15	2	17
LC	2	6	8
Total	17	8	

Table 1. Number of subjects of each classified type by actual brain condition.

Table 2
Marginal Posterior Distributions

	LC Heuristic		VM Heuristic	
	Mean	SD	Mean	SD
Constant	-0.05362	0.09747	0.04513	0.08600
Loss Bad Deck	0.00062	0.00028	0.00066	0.00033
Loss Good Deck	-0.00381	0.00138	-0.00400	0.00125

Note. These coefficients correspond to the differenced value function described in section 4.b, where the bad deck is deck “1” and the good deck is deck “2.” Hence, the constant is $b_{11}-b_{12}$, Loss Bad Deck is b_{21} , and Loss Good Deck is $-b_{22}$.

**Mean Number of Bad Deck Choices by
Frontals and Controls**

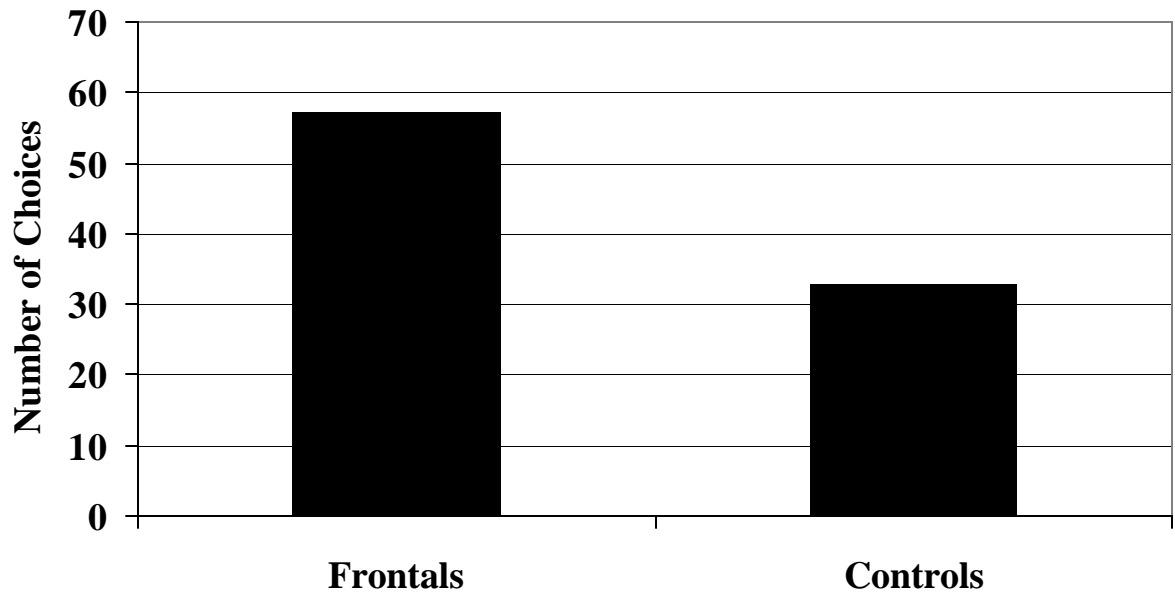


Figure 1.

Cumulative Number of "Bad Deck" Choices by Frontals and Controls

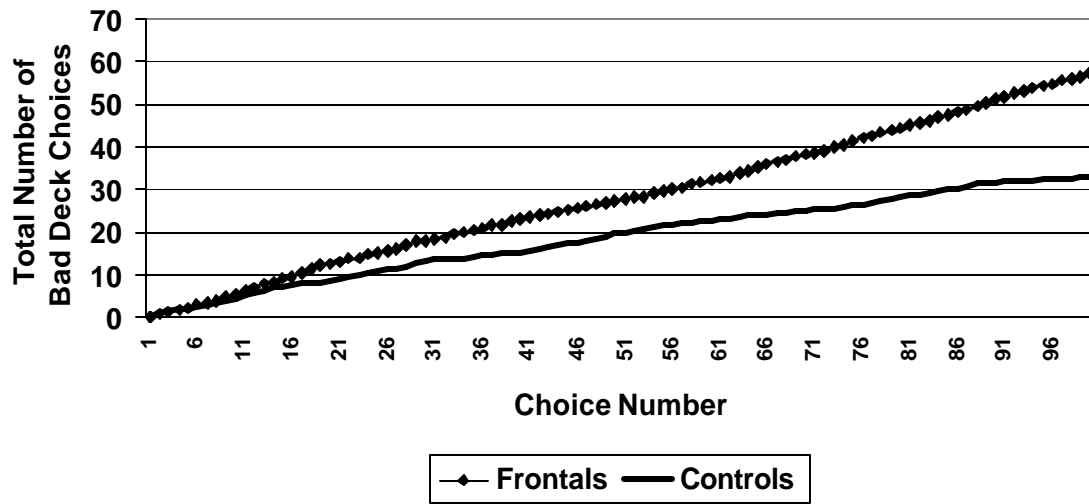


Figure 2.