Do humans perceive temporal order in asset returns?

Jasmina Hasanhodzic,* **Andrew W. Lo,†** **and Emanuele Viola‡**

December 17, 2011

Abstract

Both policymakers and academics have recently been studying whether graphical, as opposed to numerical, presentation of financial information to investors affects investors' decisions. Of particular interest is the usefulness of graphical representations of assets' performance via temporal charts of asset prices. Such charts are prevalent in financial media and investment disclosures, and are studied routinely, with the naked eye, by both casual and professional investors. This brings to the forefront a fundamental question: just what information can human beings extract from such charts? Although some anecdotal evidence has suggested that humans cannot even distinguish price charts from charts generated via a synthetic random walk, to our knowledge the above question has not previously been addressed scientifically.

We make a step towards answering this question by running an experiment to test whether human subjects can differentiate between actual vs. randomized asset returns. Our experiment consists of an online video game (`http://arora.ccs.neu.edu`) where players are challenged to distinguish actual price charts from "randomized" price charts obtained by randomly permuting the actual returns, or price differences. Implementing the experiment via a video game allows us to collect a large amount of data efficiently, while making the process fun for the subjects, so that they do not get tired and distort their behavior. In all but one of our eight datasets, we find statistical evidence that subjects can distinguish between actual and randomized price charts.

These results show that temporal charts of asset prices convey to investors information that cannot be reproduced by summary statistics, and call for more research on the usefulness of making such representations available to investors. Our results contrast previous anecdotal evidence.

We then compare subjects' performance with autocorrelations of various moments of the data. Finally, we group subjects according to various demographic categories, and analyze their relative performance. In both cases our results reveal interesting relationships.

*Graduate student, Boston University Department of Economics. Corresponding author: Jasmina Hasanhodzic (646) 338–1971 (voice), `jah@bu.edu` (email).

†Harris & Harris Group Professor, MIT Sloan School of Management, Cambridge, MA 02142, (617) 253–0920 (voice), `alo@mit.edu` (email), and Chief Investment Strategist, AlphaSimplex Group. Research support from the MIT Laboratory for Financial Engineering is gratefully acknowledged.

‡Assistant Professor, Northeastern University, College of Computer and Information Science, 360 Huntington Avenue, Boston, MA 02115, (617) 373–8298 (voice), `viola@ccs.neu.edu` (email). Partially supported by NSF grant CCF–0845003.

1 Introduction

One of the most important and complex decisions individuals face is how to save and invest. Choices they make affect not only their own quality of life, but may have an impact on the economy by creating dependencies on government-sponsored benefits. However, it is well noted that when it comes to investing, individuals are not well positioned to make sound decisions. Several reasons have been proposed in the literature, ranging from overload of information about investment products to choose from, marketing strategies designed to mislead, behavioral biases, and financial illiteracy; see, for example, Bazerman (2001), Bodie (2007), Choi, Laibson, and Madrian (2010), and the references therein. The problem of inadequate individual investment decisions is especially acute in the case of retirement savings, where the recent shift from defined benefit pension plans to privatized 401(k) plans has forced individuals to, in effect, manage their own money. As a result, much debate among policymakers and academics has taken place about improving the quality and presentation of data available to investors. For example Bazerman (2001) and Kozup et al. (2008) call for research on investors perceptions of investment products, and ways of making the information about those products easy to access and comprehend.

An example of recent academic work in this direction is Hung, Heinberg, and Yoong (2010), who evaluate versions of the Department of Labor's proposed Model Comparative Chart, which provides a standard simplified disclosure format for investment information. They conduct an online experiment where subjects are asked to allocate $10,000 among different funds based on funds performance disclosure. In one version of the disclosure, past returns are presented as a numerical table. In another version, in addition to the numerical table, the disclosure shows a graphical representation of returns over a 10-year period, as a bar chart. For completeness, a relevant figure from their work is reproduced in Figure A.1 in the Appendix at the end of this paper. Perhaps surprisingly, the authors find that the two disclosures have a statistically significant effect on the retirement investment allocation, although the effect may not be practically significant

in terms of investment outcomes.

Together with the prevalence of temporal charts of asset returns in financial media such as Yahoo! Finance, and their widespread use by both casual and professional investors, the above brings to the forefront a fundamental question: Just what information can human beings extract from charts of financial returns? This question has several ramifications. For example: Are there any patterns in financial asset returns that humans can actually extract by looking at such charts? Is seeing a chart more informative than just having a few parameters like, say, average and variance? Could Yahoo! and numerous other websites that display charts save space by getting rid of them altogether, with no harm to investors? Would Hung, Heinberg, and Yoong's (2010) experiment have the same outcome if subjects were presented with a "random" chart? In other words, is the mere *presence of some chart* of the data biasing the subjects, or are subjects actually gathering information from the *contents of the chart*?

In this paper we make a step towards answering these questions by addressing a natural question on the ability of human beings to extract information from charts of financial data:

do humans perceive temporal order in asset returns? ($\star$)

This question is of interest in light of the aforementioned debate. If the answer is negative, then temporal charts of asset returns can be safely replaced by summary statistics that are oblivious to temporal dependencies. If positive, such charts may be useful to investors' decisions.

In this paper we report the results of an experiment designed to test the ability of human subjects to distinguish between actual and randomly generated charts of financial asset prices. We develop a simple web-based video-game, available at `http://arora.ccs.neu.edu`. In this game, subjects are shown two dynamic price series (i.e. moving charts) side by side—both of which display price graphs evolving in real time (a new price realized roughly each second)—but only one of which is a "replay" of actual historical price series. The other series is constructed from a random shuffling of the actual series, which preserves

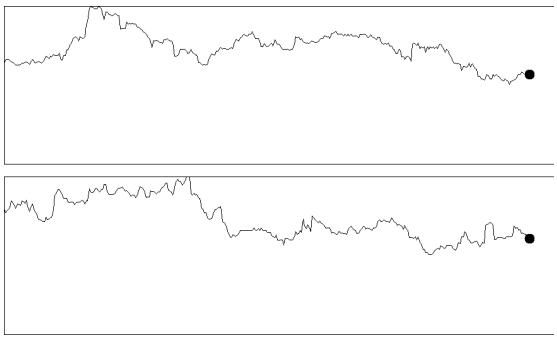


Figure 1: Reindeer (real data in top panel).

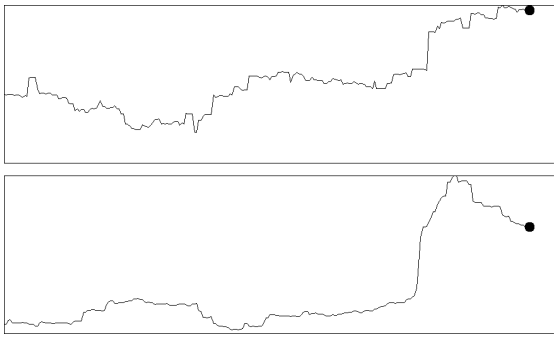


Figure 2: Bear (real data in bottom panel).

Figure 3: Wrong choice in Beaver contest.

Figure 4: Correct choice in Elk contest.

the marginal distribution of the returns but eliminates any time-series properties, see Figures 1 and 2. Subjects are asked to press a button indicating their selection of the actual price series, and are informed immediately whether they were correct or incorrect, see Figures 3 and 4, after which the next pair of price series begins being displayed. Note that the charts are moving, so at any point in time there is a certain number of observations present on the screen for each time series, which is a subset of the total number of observations subjects see on a moving chart before having to make a guess (these parameters are reported for each dataset later in the paper). Subjects do not have to wait until the entire moving chart is completed being displayed before making their choice, but can guess at any time prior to its completion (an omnipresent counter informs them of the time left). They have a counter telling them how many seconds they have remaining before the moving chart is done. The game is fast-paced: subjects can observe the charts for 10 to 25 seconds (depending on the dataset) before having to make a guess.[1]

We note here that to obtain a random chart, we permute asset returns but show the subjects their corresponding prices. By contrast, Hung, Heinberg, and Yoong (2010) show the subjects charts of returns. Our choice is motivated by the fact that price charts are routinely used by both major and casual investors, and are widely disseminated through the news media and Internet sites such as *Yahoo! Finance*. Even those without any investment experience are exposed to such charts. Hence understanding price charts seems a more natural first step.

In a sample of 78 subjects participating in up to 8 different contests (using different types of financial data),[2] with each contest lasting two weeks and concluding with prizes awarded to top performers, we obtained 8015 human-generated guesses for this real-time choice problem. For all but one of our eight datasets, the results provide strong statistical evidence that humans can distinguish actual price series from randomly generated ones.

We stress that the idea of testing the ability of human subjects to distinguish random vs. real

[1] Jumping ahead, we used 8 different datasets named after animals. Subjects are given 11 seconds to guess in the Bull contest, 15 in Bear, Elk, and Raindeer, 20 in Lynx and Mandrill, 22 in Seal, and 24 in Beaver.

[2] Specifically, 78 accounts were created, each corresponding to a unique e-mail address.

data using graphical representations is not new. Indeed, this has been studied in depth in the Information Visualization literature, see for example the works by Heer, Kong, and Agrawala (2009), and Wickham, Cook, Hofmann, and Buja (2010), and the references therein. However, we are unaware of any previous work where this idea has been used in a financial setting.

Similarly, we do not view the video game we developed as a main contribution of this paper. This video game is not intended to represent a novel tool for visualizing financial data. In fact, it displays data in a fashion similar to commonly used trading platforms; and similar tools are for example reviewed in the Information Visualization papers just cited.

Instead, implementing the experiment as a video game is intended to make the process fun and engaging for the subjects, so that they do not get tired, bored, or frustrated in a way that might affect their behavior. Moreover, the game allows the subjects to make their choices quickly, allowing us to get a large amount of data efficiently, with as little cost to subjects as possible, with the goal of answering the null hypothesis effectively.

The main contributions of this paper are to address a fundamental finance question that recent developments have brought to the forefront but that was never analyzed scientifically, and to draw conclusions which contrast previous anecdotal evidence. An additional contribution of this paper is the application of video-game-like interfaces to a financial setting.

Indeed, until the present paper, the anecdotal evidence has been that the answer to the question ($\star$) is negative. Specifically, it was argued that humans cannot tell price charts from "random," such as charts generated by a random walk. For example, in an experiment (Malkiel 1973, p. 143) students were asked to generate returns (i.e., price differences) by tossing fair coins, and it was argued that those yielded observations that were indistinguishable from market returns to human subjects observing corresponding price charts. For similar arguments in the finance literature see, for example, Roberts (1959), Kroll, Levy, and Rapoport (1988), DeBondt (1993), Wärneryd (2001), and Swedroe (2005). Such anecdotal evidence has also been collected in the computer science literature. For example, Keogh and Kasetty (2003) "asked 12 professors at UCRs Anderson Graduate

School of Management to look at Figure A.2 (included at the end of this paper) and determine which three sequences are random walk, and which three are real S&P500 stocks." They find that "the accuracy of the humans was 55.6%, which does not differ significantly from random guessing."

Although this anecdotal evidence has suggested that humans cannot distinguish price charts from charts generated via a synthetic random walk, to our knowledge this question has not previously been addressed using scientific methods. Note that for charts generated by random walks it is statistically impossible to perform well in our experiment. On the other hand, we demonstrate that human beings perform well for charts of financial assets. Hence, our results contrast the anecdotal evidence.

Our results are also of interest in light of the hotly debated Efficient Market Hypothesis according to which "prices fully reflect all available information" and hence must be unforecastable; see, for example, Samuelson (1965), Fama (1965a), Fama (1965b), and Fama (1970). More recent works, such as Lo and MacKinlay (1988, 1999) and Lo, Mamayski, and Wang (2000), provide compelling evidence that markets are not efficient, i.e. price data does possess statistical properties that noticeably deviate from random models. In fact, autocorrelation is such a property. However, we point out that the data analysis in all these works is *computer*, not *human*-based. Consequently, the works leave open the fundamental question of whether markets look efficient *to human beings*. Our work appears to be the first to provide evidence that, in fact, even human beings can extract information from financial-return data in a way that would not be possible if this data was generated by a random walk.

We also would like to mention that the results presented in this paper comprise our entire experiment. Moreover, our experiment is easily reproducible: skeptics may wish to try the challenge for themselves, and demonstrate that in short order, they can become quite skilled at differentiating real financial data from randomized series.

This paper is organized as follows. In Section 2 we describe our experiment. In Section 3 we

describe our results, including their breakdown by demographic group. In Section 4 we discuss properties of our datasets. Finally, we conclude in Section 5.

2 Experiment Design

To test the null hypothesis H that human subjects cannot distinguish between actual and randomly generated price series, we begin with a time series of actual historical prices $\{p_0, p_1, p_2, \ldots, p_T\}$ and compute the returns or price differences $\{r_t\}$,

$$r_t \;\equiv\; p_t \;-\; p_{t-1}.$$

From this, we construct a randomly generated price series $\{p_0^*, p_2^*, \ldots, p_T^*\}$ by cumulating randomly permuted returns:

$$\begin{aligned} p_t^* \;&\equiv\; \sum_{k=1}^{t} r_{\pi(k)} \quad , \quad p_0^* \;\equiv\; p_0 \quad , \\ \pi(k) \;&:\; \{1, \ldots, T\} \;\rightarrow\; \{1, \ldots, T\} \end{aligned}$$

where $\pi(k)$ is a uniform permutation of the set of time indexes $\{1, \ldots, T\}$. A random permutation of the actual returns does not alter the marginal distribution of the returns, but it does destroy the time-series structure of the original series, including any temporal patterns contained in the data. Therefore, the randomly permuted returns will have the same mean, standard deviation, and moments of higher order as the actual return series, but will not contain any time-series patterns that may be used for prediction. This construction will allow us to test specifically for the ability of human subjects to detect temporal dependencies in financial data.

To implement this comparison, we developed a web-based video-game (`http://arora.ccs.neu.edu`), which was advertised via email and on websites to com-

puter science and finance students at our institutions, as well as to a number of associations for individual investors, see Figure A.3 in the Appendix for a sample advertisement.[3]

To register, a subject has to fill a short demographic questionnaire where they were asked to select one of prespecified categories describing occupation (academic, finance, student, other), sex (male, female), education (high school, undergraduate, MS, PhD), age, and country.

After registration, a subject can participate in trials from eight different contests, each consisting of the same game applied to different datasets. The datasets consist of returns of eight commonly traded financial assets: the NASDAQ Composite Index, the Russell 2000 Index, the US Dollar Index, Gold (spot price), the Dow Jones Corporate Bond Price Index, the Dow Jones Industrial Average, the Canada/US Dollar Foreign Exchange Rate, and the S&P GSCI Corn Index (spot price).[4] These datasets were arbitrarily named after animals, so that users had no knowledge of the specific financial assets used in the experiment.

Participating in a trial consists of the following task. The subject is shown two dynamic price charts on a computer screen, one above the other (Figures 1 and 2). Each graph evolves through time—similar to those appearing in computer trading platforms—plotting the price at that point in time as well as the trailing prices over a fixed time window over the most recent past. Prices are defined as the cumulative sum of a sequence of returns. Of the two moving charts, only one corresponds to the sequence of market returns from the actual dataset; we call this graph the "real" chart or $\{p_t\}$. The other corresponds to the sequence of returns obtained by randomly permuting the sequence of market returns in the real chart; we call this graph the "random chart" or $\{p_t^*\}$. The computer chooses at random which of the two graphs is placed at the top or the bottom.

The subject is asked to decide which of the two moving charts is the real one by clicking on

[3]To recruit subjects, an announcement was emailed to Northeastern computer science students, MIT Sloan MBA students in the Fall section of 15.970, members of the American Association of Individual Investors mailing list, Market Technicians Association mailing list, the MTA Educational Foundation mailing list, and the staff and Twitter followers of TraderPsyches.

[4]The Dow Jones Corporate Bond Price Index was obtained from the Global Financial Database, while all other data series were obtained from Bloomberg.

it. The game registers the subject's choice, and informs the subject immediately whether his/her guess is correct or incorrect, see Figures 3 and 4. For each dataset, the user is shown approximately 35 pairs of moving charts and asked to make as many choices. The subject is also free to refrain from choosing. This happened rarely, and to err on the conservative side, we recorded the absence of a guess as an incorrect choice for that trial. To provide the participants with some incentive for making correct choices, top-scoring players were awarded prizes ($10 or $25 Amazon gift certificates).

To evaluate the robustness of our experimental design, we varied various parameters of the experiment across datasets, as indicated in the Results section below. In addition, we presented subjects with data charts in two different ways. For half of the datasets corresponding to transaction-by-transaction (or "tick") data, each subject was shown a fresh set of charts, based on a sequence of returns disjoint from the sequences shown to any other subjects. For the other half of the data, corresponding to daily data, the charts shown to each subject were based on the same sequence of returns.[5]

Finally, for each dataset, subjects were offered the opportunity to train on a disjoint set of data.

3 Results

The results are summarized in Figures 5–7. For each dataset we report how many return observations are presented to each subject (points per chart). As the charts are moving, we also report how many returns are present on the screen at any moment (points per screen). We then report how many pairs of charts each subject was presented with (charts per subject), and how many subjects participated (subjects). The distribution of correct guesses across subjects is reported in a histogram, together with the corresponding p-value, computed using the Student's t-test. For all but one of our eight datasets, we obtain statistical evidence (p-value less than 5%) that humans can

[5]However, the data was shifted by a random amount for security reasons, i.e., to avoid the possibility that two subjects could coordinate their guesses, for example by simultaneously playing the same charts on two nearby machines.

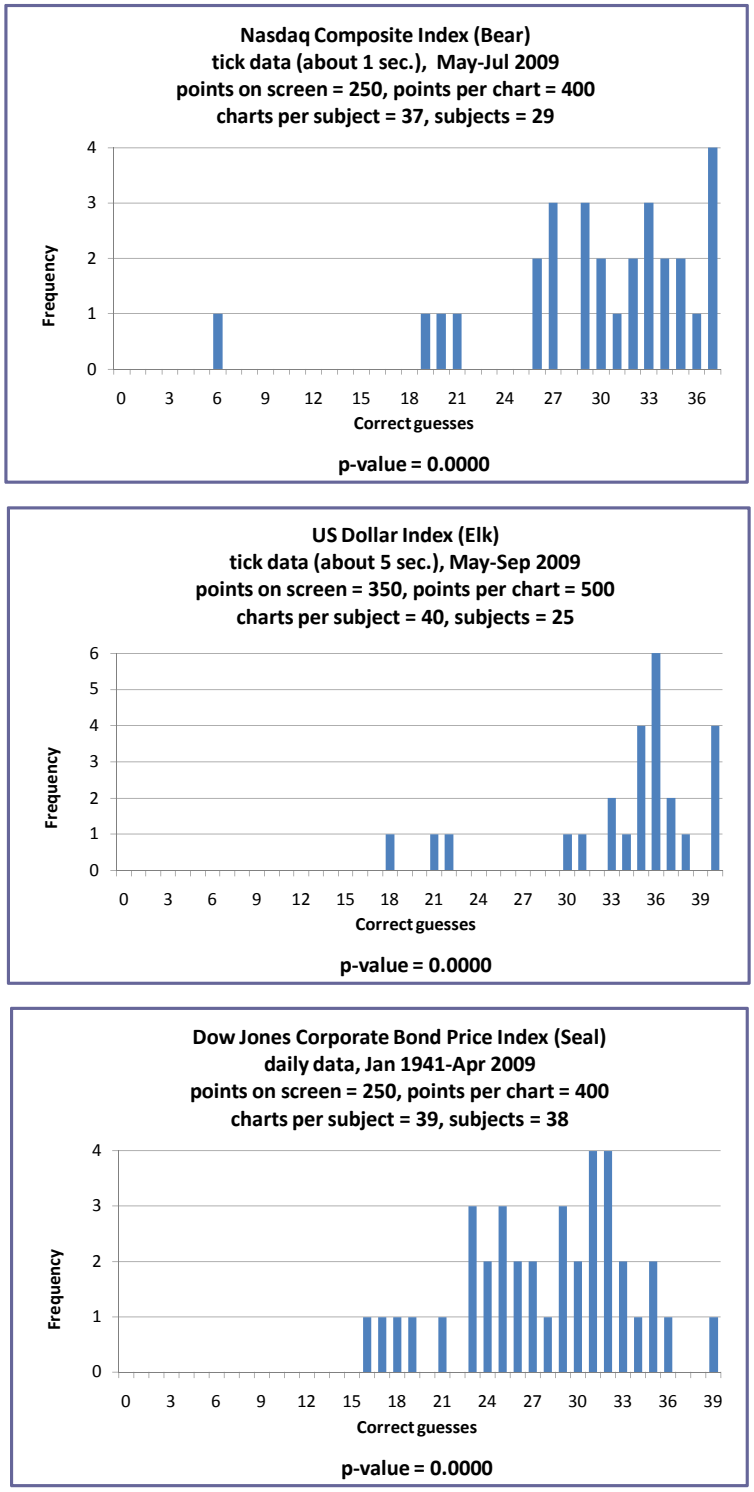


Figure 5: Histograms of experimental results for Nasdaq Composite Index (Bear), US Dollar Index (Elk), and Dow Jones Corporate Bond Price Index (Seal) contests. For each histogram, we report data description and experiment parameters in the title bar, and p-values at the bottom of the chart.

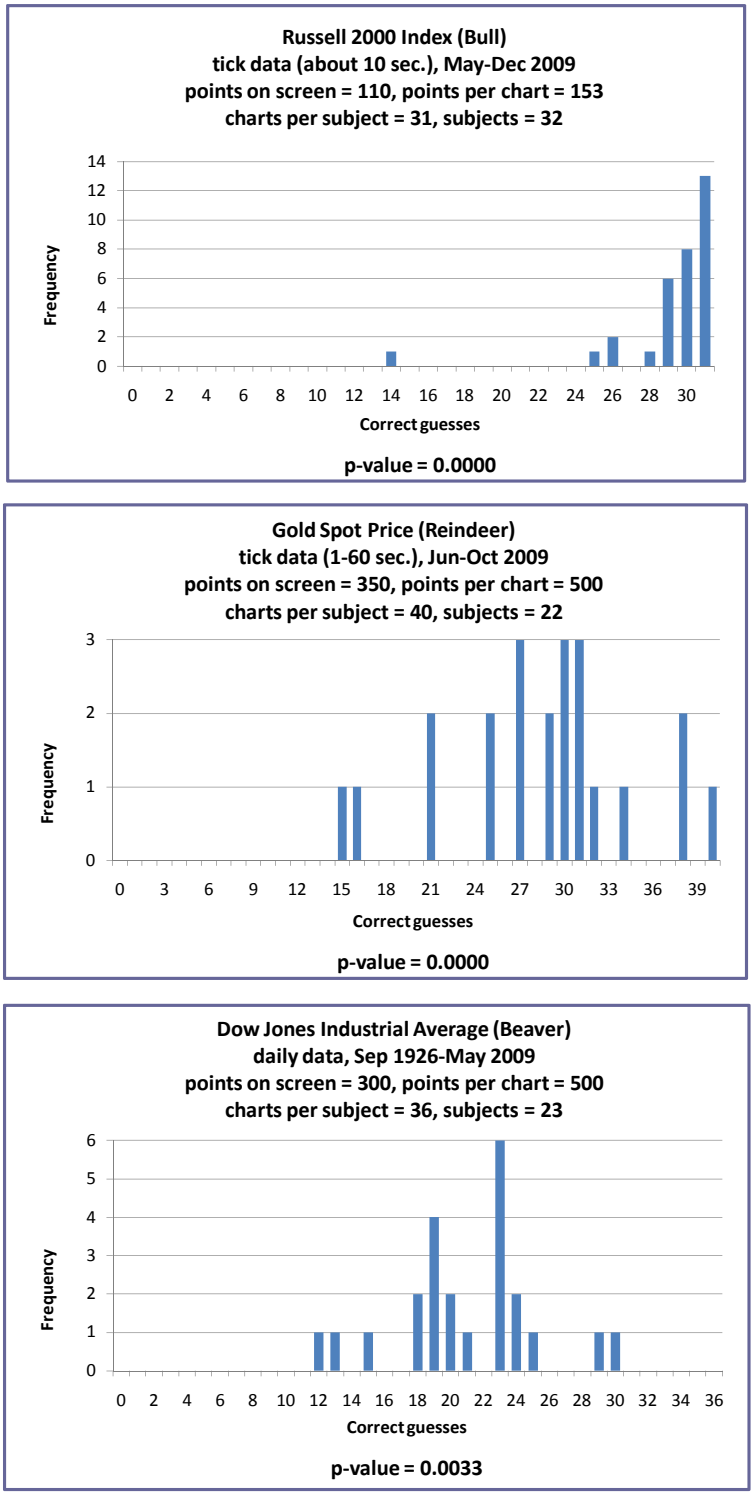


Figure 6: Histograms of experimental results for Russell 2000 Index (Bull), Gold Spot Price (Reindeer), and Dow Jones Industrial Average (Beaver) contests. For each histogram, we report data description and experiment parameters in the title bar, and p-values at the bottom of the chart.

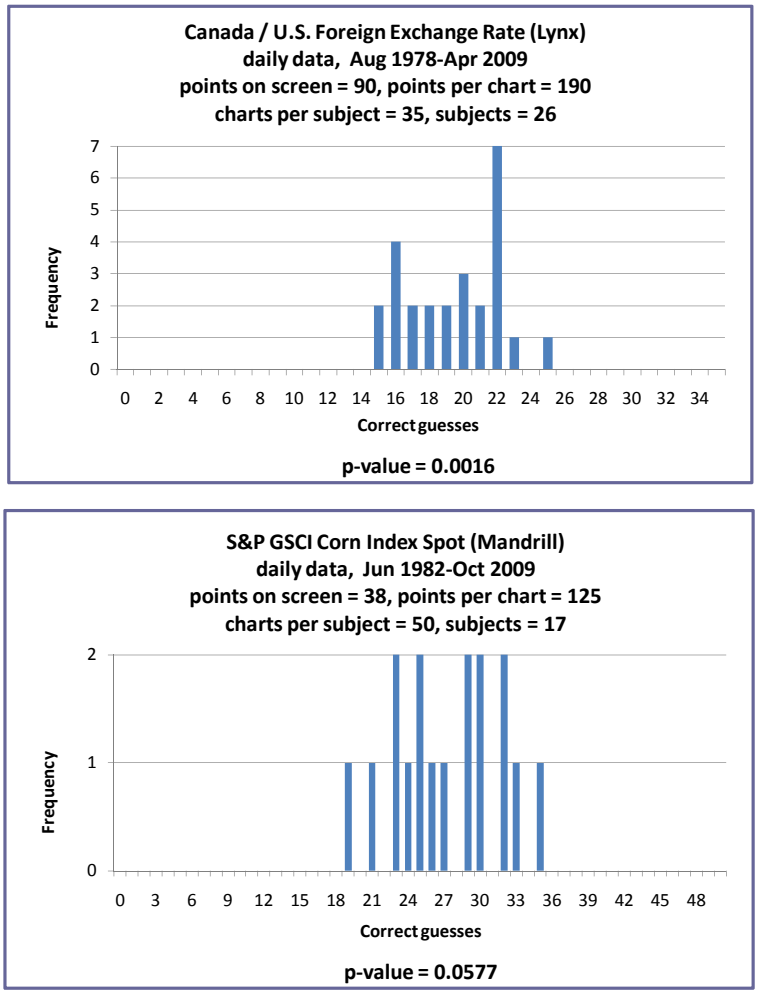


Figure 7: Histograms of experimental results for Canada/U.S. Foreign Exchange Rate (Lynx) and S&P GSCI Corn Index Spot (Mandrill) contests. For each histogram, we report data description and experiment parameters in the title bar, and p-values at the bottom of the chart.

tell the two types of charts apart.

The case in which we do not obtain statistical evidence at 5% level is the one corresponding to S&P GSCI Corn Index (spot) dataset, coded as Mandrill. An obvious difference between this contest and the others is this: While playing Mandrill, subjects are asked to make their guess based on at most 38 daily observations per screen, whereas for other contests this value is at least twice as high and in most cases an order of magnitude higher. Hence our results may suggest that with too few observations there is too much randomness in the data for temporal order to emerge in the eyes of the subjects. We note however that even for this dataset the p-value is less than 6%.

3.1 Demographic breakdown

In this section we discuss how well subjects from different demographic groups performed. The results are summarized in Tables 1–3. To avoid clutter, p-values are reported in percentage points, unlike in Figures 5–7. Specifically, in Table 1 we divide subjects in groups based on occupation (academic or other, finance, student), sex (female, male), education (high school or undergraduate, MS or PhD), age (30 and above, below 30), and country (USA, rest of the world). Recall from Section 2 that each subject had to fill a short demographic questionnaire where they were asked to select one of prespecified categories describing occupation, sex, education, age, and country. Some categories had too few subjects, so we combined them into one group. For one example regarding occupation, academic and other are combined in the group acamidec/other, see Table 1.

For each group we report the Student's t-test p-values of the distribution of correct guesses across subjects for the eight contests under consideration. We also report the number of subjects participating in each contest. For reference, we report again the corresponding values for the entire sample, previously reported in Figures 5–7. We highlight in red the p-values greater than 5%. For example, the 6 subjects who indicated "academic" or "other" for occupation had a p-value of 3.5% when playing the Gold Spot Price (Reindeer) contest.

To compare relative performance across groups we compare the corresponding p-values. One interesting pattern that emerges is that males outperform females across the board. And the group of males even rejects the null hypothesis at the 5% level for the S&P GSCI Corn Index Spot (Mandrill) contest. Also, younger subjects do significantly better than the older ones in three contests, and marginally worse in one. This may have to do with the fact that these groups—males and younger subjects—may be more used to the video-game kind of interface that was used in our experiment. The data also suggests that subjects with finance background do not perform better than others. For example, student subjects beat finance subjects in four contests, and lose in one.

In Tables 2–3 we summarize each group's performance across all contests. We do this in two different ways to confirm the robustness of the results. In Table 2 we again use Student's t-test, but we drop the Russell 2000 Index (Bull) and S&P GSCI Corn Index Spot (Mandrill) contests because in these two contests the number of guesses each subject had to make deviates substantially from that of the other contests. (Bull and Mandrill's "charts per subject" parameters are 31 and 50, respectively, vs. 35 to 40 for the other contests.) So, after dropping Bull and Mandrill, the number of guesses each subject had to make is roughly on the same scale across contests. Another reason to drop Bull and Mandrill is that those are the contests where subjects performed best and worst, respectively, so they may not be informative.

As our second way to summarize each group's performance across all contests, we consider all subjects from a particular demographic group as a single person. We add up the correct guesses across all subjects in that group (call this sum g), as well as the corresponding total guesses (call this sum n). Under the null hypothesis H, human subjects should not be able to distinguish between real and random charts, so their choices should be no better than purely random guesses. Therefore, testing the null hypothesis involves computing the p-value of obtaining at least as many correct guesses when guessing at random, i.e., by tossing a fair coin. Hence, in this case the p-value is simply computed as the probability mass of the tail of the binomial distributions, that is, as the

probability that the number X of "heads" in n independent tosses of a fair coin is at least g:

$$p\text{-value} \equiv \Pr[X \geq g] = \sum_{i=g}^{n} \binom{n}{i} / 2^n \,. \tag{1}$$

The results are similar to those obtained with the previous method. They are reported in Table 3.

A striking fact emerges regardless of the approach: academics do poorly. In combination with our previous observation that better-performing subjects tend to be younger and to not have a finance background, and with the fact that the experiment was relatively fast paced, these results suggest that in our experiment it was advantageous to take an intuitive as opposed to analytical approach.

4 Data properties

To gain some insight into the essential properties of the data that subjects may have been calibrating on, in Tables 4–5, for each of the contest datasets and their randomized counterparts, we consider the time-series properties of returns, as well as of the higher-order moments of returns (squared, cubed, and fourth power returns). In each case, we compute first order autocorrelation coefficients and p-values of the Ljung-Box Q-statistics, based on the first through 20th autocorrelation coefficients. The Ljung-Box Q-statistics provide a measure of the statistical significance of autocorrelation in the returns, with smaller p-values indicating more statistically significant autocorrelations. These statistics are computed chart by chart, that is, over nonoverlapping windows of length given by the parameter "points per chart", corresponding to the data windows on which subjects needed to make a guess. We compute mean, standard deviation, and percentiles across these chart-by-chart autocorrelations and p-values of Q_{20}-statistics. Note that the total number of autocorrelation coefficients (or equivalently p-values) that we are averaging over equals "charts per

Demographic group		# subjects	*p*-value (%)	# subjects	*p*-value (%)	# subjects	*p*-value (%)	# subjects	*p*-value (%)
		Bear		**Elk**		**Bull**		**Reindeer**	
Occup	academic/other	9	0.0	8	0.2	7	0.2	6	3.5
	finance	9	0.0	7	0.0	10	0.0	9	0.2
	student	11	0.0	10	0.0	15	0.0	7	2.6
Sex	female	5	**19.9**	7	0.1	6	0.0	4	0.1
	male	24	0.0	18	0.0	26	0.0	18	0.0
Educ	high school/undergrad	12	0.0	8	0.0	14	0.0	8	0.8
	MS/PhD	17	0.0	17	0.0	18	0.0	14	0.0
Age	>=30	17	0.0	16	0.0	19	0.0	13	0.0
	<30	12	0.0	9	0.0	13	0.0	9	1.1
Country	USA	17	1.6	15	0.0	18	0.0	13	0.3
	other	12	0.0	10	0.0	14	0.0	9	0.0
Entire sample		29	0.0	25	0.0	32	0.0	22	0.0
		Beaver		**Seal**		**Lynx**		**Mandrill**	
Occup	academic/other	9	**27.7**	9	0.5	7	3.5	5	**60.3**
	finance	6	**14.1**	7	0.6	7	**31.3**	4	**27.8**
	student	8	0.5	22	0.0	12	4.9	8	**10.9**
Sex	female	8	**6.8**	8	2.2	7	4.7	7	**93.3**
	male	15	2.9	30	0.0	19	1.9	10	2.2
Educ	high school/undergrad	10	3.7	17	0.0	13	**7.8**	8	**22.7**
	MS/PhD	13	4.4	21	0.0	13	1.0	9	**17.9**
Age	>=30	13	**14.3**	14	0.0	12	**5.5**	7	**53.2**
	<30	10	0.1	24	0.0	14	1.5	10	**6.7**
Country	USA	14	**27.1**	25	0.0	17	0.7	10	**48.8**
	other	9	0.2	13	0.0	9	**12.2**	7	**5.3**
Entire sample		23	0.3	38	0.0	26	0.2	17	**5.8**

Table 1: Student's t-test p-values by demographic group for Nasdaq Composite Index (Bear), US Dollar Index (Elk), Russell 2000 Index (Bull), Gold Spot Price (Reindeer), Dow Jones Industrial Average (Beaver), Dow Jones Corporate Bond Price Index (Seal), Canada/U.S. Foreign Exchange Rate (Lynx), and S&P GSCI Corn Index Spot (Mandrill) contests. The p-values greater than 5% are highlighted in red.

Demographic group		# subjects	*p*-value (%)
Occup	academic	11	**43.2**
	finance	45	0.0
	other	37	0.0
	student	70	0.0
Sex	female	39	0.0
	male	124	0.0
Educ	high school	6	0.3
	undergrad	62	0.0
	MS	69	0.0
	PhD	26	0.0
Age	[18,27)	60	0.0
	[27,45)	56	0.0
	[45,67]	47	0.0
Country	USA	101	0.0
	other	62	0.0
Entire sample		163	0.0

Table 2: Summary of each group's performance across all contests except Russell 2000 Index (Bull) or S&P GSCI Corn Index Spot (Mandrill). The p-values are computed using Student's t-test. Those greater than 5% are highlighted in red.

Demographic group		# subjects	*p*-value (%)
Occup	academic	13	**7.8**
	finance	59	0.0
	other	47	0.0
	student	93	0.0
Sex	female	52	0.0
	male	160	0.0
Educ	high school	8	0.0
	undergrad	82	0.0
	MS	88	0.0
	PhD	34	0.0
Age	[18,27)	79	0.0
	[27,45)	73	0.0
	[45,67]	60	0.0
Country	USA	129	0.0
	other	83	0.0
Entire sample		212	0.0

Table 3: Summary of each group's performance across all contests. Each group is considered as a single individual. The p-values are computed as the probability mass of the tail of the binomial distribution, cf. Equation 1. Those greater than 5% are highlighted in red.

subject" for the daily datasets and "charts per subject" times "subjects" for the tick datasets (recall that in the latter case, each subject was shown a fresh set of charts, based on data disjoint from that used with other subjects). Autocorrelations of returns, and of squared, cubed, and fourth power returns, indicate the persistence in the return sign, volatility, skewness, and kurtosis, respectively, all of which are properties of the data that subjects could have been using to distinguish actual from randomized series.

For the randomized data for all datasets and all moments of returns under consideration, the average autocorrelation coefficients are close to zero, and the average p-values of the Q_{20}-statistics are high, ranging from 47% for the Dow Jones Industrial Average (Beaver) to 88% for the Nasdaq Composite Index (Bear). This is due to the fact that the shuffling process destroys time-series dependencies. Subjects performed best in the contests with lowest p-values (Q_{20}): the actual returns in each of Dow Jones Corporate Bond Price Index (Seal), Nasdaq Composite Index (Bear), US Dollar Index (Elk), and Russell 2000 Index (Bull) datasets have average p-values (Q_{20}) of 12%, 7%, 5%, and 7%, respectively, which is significantly lower than the corresponding values for their randomized counterparts (51%, 60%, 51%, and 61%). For Dow Jones Industrial Average (Beaver), Canada/US Foreign Exchnage Rate (Lynx), and Gold Spot Price (Reindeer) datasets, where subjects still performed well enough to reject the null hypothesis, the difference in average p-values (Q_{20}) between actual and randomized returns is somewhat smaller (19%, 31%, and 19% for the actual versus 47%, 57%, and 53% for the randomized returns, respectively).[6] The difference in average p-values (Q_{20}) is the smallest for the case of S&P GSCI Corn Index Spot (Mandrill), where the null hypothesis is not rejected (40% for the actual returns versus 55% for their randomized counterpart).

While the significance of autocorrelations of returns seems to be a good indicator of the subjects' performance, it is not the only property of the data that subjects could have been calibrating

[6]We note this difference is still statistically significant at the 5% level. This is computed by taking the corresponding standard deviations into account and using the differences-in-means t-test.

on. The differences in average p-values (Q_{20}) between actual and randomized squared and cubed returns are also statistically significant—for example, 16% vs. 58% for squared returns and 36% vs. 78% for cubed returns for Dow Jones Industrial Average (Beaver), 18% vs. 63% for squared returns and 30% vs. 71% for cubed returns for Dow Jones Corporate Bond Price Index (Seal), 20% vs. 60% for squared returns and 32% vs. 70% for cubed returns for the US Dollar Index (Elk), and 22% vs. 61% for squared returns and 35% vs. 72% for cubed returns for Gold Spot Price (Reindeer). This suggests that subjects may have taken higher order moments into account when formulating their guesses. Whether humans are able to perform well using higher-order moments and more complex patterns than simply linear forecasting rules, which is what autocorrelation of returns captures, deserves further scrutiny and is left for future research.

Similar autocorrelation patterns are observed in Tables A.1–A.2 in the appendix, corresponding to the portions of the datasets on which subjects were allowed to train before entering the contest. This indicates that if, indeed, subjects were calibrating on the properties of the data described above, the training portions of the datasets provided solid training ground. This may explain how some subjects were able to have all of their guesses correct, as in the case of Russell 2000 Index (Bull) contest.

5 Conclusion

A natural question that arises is how were the subjects able to perform so well in seven out of eight datasets?

Several facts suggest that subjects have taken an intuitive as opposed to analytical approach. One fact is that the typically more analytical group, academics, performed poorly. Another is that subjects with a finance background performed poorly, so understanding financial data was not a plus. Finally, the game is relatively fast paced and so subjects have little time to think about their guesses.

Ret type	Data type	Mean	Std	1%	25%	50%	75%	99%	Mean	Std	1%	25%	50%	75%	99%
		1st order autocorrelation (%)							*p*-value of Ljung-Box Q_{20} (%)						
		Nasdaq Composite Index (Bear)													
r	**actual**	**25**	11	-1	17	25	33	51	**7**	20	0	0	0	0	100
	rand	**0**	5	-12	-3	0	2	13	**60**	31	0	35	66	88	100
r^2	**actual**	**12**	11	-2	3	8	17	48	**49**	44	0	0	49	99	100
	rand	**0**	4	-5	-2	-1	0	19	**78**	35	0	70	99	100	100
r^3	**actual**	**7**	10	-2	1	3	10	46	**66**	43	0	10	99	100	100
	rand	**0**	3	-9	-1	0	0	9	**83**	33	0	94	100	100	100
r^4	**actual**	**4**	9	-1	0	0	5	44	**77**	39	0	67	100	100	100
	rand	**0**	3	-2	-1	-1	0	19	**88**	30	0	100	100	100	100
		US Dollar Index (Elk)													
r	**actual**	**-23**	9	-41	-29	-24	-18	4	**5**	15	0	0	0	1	90
	rand	**0**	4	-10	-3	0	3	11	**51**	30	1	24	52	77	99
r^2	**actual**	**13**	9	-1	7	12	18	39	**20**	30	0	0	3	28	100
	rand	**0**	4	-9	-3	-1	2	14	**60**	32	0	33	66	90	100
r^3	**actual**	**-11**	11	-43	-17	-10	-4	16	**32**	38	0	0	7	66	100
	rand	**0**	4	-11	-2	0	1	11	**70**	35	0	47	87	99	100
r^4	**actual**	**8**	10	-2	1	5	11	45	**51**	43	0	1	56	99	100
	rand	**0**	3	-5	-2	-1	0	14	**79**	34	0	70	99	100	100
		Russell 2000 Index (Bull)													
r	**actual**	**42**	15	4	34	44	53	71	**7**	21	0	0	0	0	100
	rand	**-1**	7	-19	-5	-1	4	17	**61**	31	1	35	67	89	100
r^2	**actual**	**15**	14	-3	3	11	23	57	**58**	42	0	10	72	100	100
	rand	**-1**	6	-9	-4	-2	0	22	**78**	31	0	65	97	100	100
r^3	**actual**	**10**	13	-2	1	6	16	54	**71**	40	0	33	99	100	100
	rand	**0**	6	-14	-2	-1	0	26	**83**	31	0	86	100	100	100
r^4	**actual**	**5**	11	-3	-1	0	7	49	**80**	35	0	80	100	100	100
	rand	**-1**	5	-6	-2	-1	-1	26	**87**	28	0	99	100	100	100
		Gold Spot Price (Reindeer)													
r	**actual**	**-18**	19	-64	-29	-13	-3	12	**19**	27	0	0	2	31	95
	rand	**0**	5	-11	-3	0	3	10	**53**	29	0	28	55	77	100
r^2	**actual**	**15**	13	-6	4	12	23	49	**22**	34	0	0	0	40	100
	rand	**0**	4	-7	-3	-1	2	13	**61**	33	0	33	68	93	100
r^3	**actual**	**-14**	18	-56	-25	-8	-1	18	**35**	42	0	0	5	88	100
	rand	**0**	4	-11	-1	0	1	12	**72**	34	0	49	89	100	100
r^4	**actual**	**9**	13	-4	0	4	16	49	**48**	46	0	0	42	100	100
	rand	**0**	3	-5	-2	-1	0	14	**79**	33	0	69	99	100	100

Table 4: Properties of contest daily datasets. For each of the daily datasets and their randomized counterparts used for contests, we compute 1st order autocorrelation coefficients, and p-values of the Ljung-Box Q_{20} statistics for various moments of the returns, over nonoverlapping windows of the data corresponding to charts on which subjects were taking a stand. We report mean, standard deviation, and percentiles of these chart-by-chart autocorrelations and p-values (Q_{20}).

Ret type	Data type	Mean	Std	1%	25%	50%	75%	99%	Mean	Std	1%	25%	50%	75%	99%
		1st order autocorrelation (%)							*p*-value of Ljung-Box Q_{20} (%)						
		Dow Jones Industrial Average (Beaver)													
r	**actual**	**7**	10	-12	2	6	12	28	**19**	27	0	0	5	29	94
	rand	**0**	4	-8	-4	1	3	10	**47**	27	4	25	46	65	98
r^2	**actual**	**9**	10	-5	3	7	13	46	**16**	28	0	0	0	21	93
	rand	**-1**	5	-7	-4	-1	0	20	**58**	36	0	22	63	95	100
r^3	**actual**	**0**	10	-30	-3	2	4	21	**36**	38	0	0	19	69	100
	rand	**0**	3	-4	-1	0	1	12	**78**	28	8	63	89	100	100
r^4	**actual**	**5**	8	-3	0	1	8	33	**50**	47	0	0	32	100	100
	rand	**-1**	3	-5	-2	-1	0	14	**80**	31	0	59	99	100	100
		Dow Jones Corporate Bond Price Index (Seal)													
r	**actual**	**10**	14	-21	1	11	18	41	**12**	23	0	0	0	14	98
	rand	**-1**	5	-11	-5	-1	3	11	**51**	29	1	30	46	77	100
r^2	**actual**	**16**	16	-2	2	12	27	49	**18**	32	0	0	0	23	100
	rand	**1**	6	-8	-3	-1	3	24	**63**	37	0	37	74	99	100
r^3	**actual**	**-4**	21	-49	-2	1	8	28	**30**	40	0	0	0	69	100
	rand	**1**	5	-9	-1	0	1	22	**71**	36	0	47	91	100	100
r^4	**actual**	**12**	16	-3	-1	4	24	48	**44**	44	0	0	31	99	100
	rand	**0**	3	-4	-2	0	1	8	**83**	32	0	87	100	100	100
		Canada/U.S. Foreign Exchange Rate (Lynx)													
r	**actual**	**3**	8	-14	-3	4	8	26	**31**	30	0	6	18	48	93
	rand	**-2**	7	-14	-7	-2	4	16	**57**	31	0	28	65	82	99
r^2	**actual**	**10**	14	-11	0	6	16	50	**29**	38	0	0	4	65	98
	rand	**1**	6	-10	-3	-1	2	23	**61**	35	0	31	69	95	100
r^3	**actual**	**1**	13	-38	-3	1	7	49	**41**	41	0	1	22	89	100
	rand	**-2**	5	-13	-6	-1	1	8	**71**	31	2	43	83	97	100
r^4	**actual**	**6**	14	-8	-1	1	9	61	**52**	44	0	2	55	99	100
	rand	**1**	6	-7	-3	-2	1	20	**71**	38	0	52	94	100	100
		S&P GSCI Corn Index Spot (Mandrill)													
r	**actual**	**7**	9	-11	0	8	12	35	**40**	33	0	9	35	71	99
	rand	**-2**	8	-23	-6	-2	3	17	**55**	31	0	25	62	82	100
r^2	**actual**	**9**	15	-12	0	4	18	59	**35**	37	0	0	18	70	100
	rand	**0**	9	-14	-8	-1	5	21	**60**	31	2	38	63	89	100
r^3	**actual**	**4**	14	-45	-1	2	12	50	**47**	37	0	7	46	80	100
	rand	**-1**	8	-25	-5	-1	2	30	**72**	30	0	49	85	99	100
r^4	**actual**	**8**	17	-6	-2	-1	10	73	**52**	44	0	3	62	99	100
	rand	**-2**	6	-12	-4	-2	-1	18	**75**	33	1	64	94	100	100

Table 5: Properties of contest tick datasets. For each of the tick datasets and their randomized counterparts used for contests, we compute 1st order autocorrelation coefficients, and p-values of the Ljung-Box Q_{20} statistics for various moments of the returns, over nonoverlapping windows of the data corresponding to charts on which subjects were taking a stand. We report mean, standard deviation, and percentiles of these chart-by-chart autocorrelations and p-values (Q_{20}).

We also conjecture that feedback—which allows subjects to learn and adapt—is a key factor in allowing typical subjects to distinguish real market returns from their randomized counterpart. Casual inspection of Figures 1–4 shows that distinguishing real data from randomized data is challenging; for some datasets the real chart tends to be smoother, as in Figure 2, while for other datasets the opposite is true, the real chart tends to be spikier, as in Figure 4. What complicates the matter further is that, as is evident from the data, the "smoothness" of actual data varies with time. Still, feedback from just a few trials seems sufficient for the user to extract characteristics of the data to be used in classifying charts in the near future.

The importance of feedback and of an intuitive approach are supported by the information about winning strategies that some of the subjects volunteered to share with us (anonymously). For example, a subject wrote:

> Admittedly, when first viewing the two datasets in the practice mode, it is impossible to tell which one is real, and which one is random, however, there is a pattern that quickly emerges and then the game becomes simple and the human eye can easily pick out the real array (often in under 1 second of time).

Having shown that human beings can distinguish price charts from charts obtained by permuting the returns at random, an important next step is to understand what properties of the data the subjects exploited. The analysis of the data properties in Section 4 shows that subjects may have been exploiting the autocorrelations of various moments of the data. Recall that we computed first to twentieth order autocorrelation of several moments of the returns. It would be extremely interesting to narrow down the orders of the autocorrelation and the moments of the data which are the most relevant. One way to do this is to run a variant of this experiment in which the data presented to the subjects is manipulated so as to eliminate autocorrelations of specific moments. For example, by randomly flipping the sign of the returns one can eliminate the autocorrelations of the first moment while preserving those of the second. We leave this for future research.

Another interesting direction is to compare humans' performance in our experiment against the performance of computers. The human eye—as opposed to a computer algorithm—may have a crucial advantage. It is well known that computers still struggle with many image-recognition and classification tasks that are trivial for humans. The same may be the case for distinguishing asset returns from randomized versions.

Given the recent regulatory push towards ensuring that "consumers have the information they need to choose the consumer financial products and services that are best for them,"[7] the study of optimal ways to present financial data to investors is of current interest. Our paper is a contribution to the growing body of literature on the usefulness of temporal charts in evaluation of financial asset performance.

Acknowledgments. We would like to thank Zvi Bodie for pointing out to us related work, Michael Coen for helpful discussions on a preliminary version of this work, especially on p-values, and François Gourio and Jawwad Noor for helpful feedback and discussions. We are also grateful to the anonymous referees for useful feedback.

References

[1] M. H. Bazerman. Consumer research for consumers. *Journal of Consumer Research*, 27(4):499–504, 2001.

[2] Z. Bodie. The challenge of investor education. In Z. Bodie, D. McLeavey, and L. B. Siegel, editors, *The Future of Life-Cycle Saving and Investing*, pages 169–171. Research Foundation of the CFA Institute, February 2008.

[3] J. Choi, D. Laibson, and B. Madrian. $100 bills on the sidewalk: Suboptimal saving in 401(k) plans. *Review of Economics and Statistics*, 2010.

[7] The Consumer Financial Protection Bureau, `http://www.consumerfinance.gov/protecting-you/`.

[4] W. P. M. De Bondt. Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of Forecasting*, 9(3):355–371, 1993.

[5] E. Fama. The behavior of stock market prices. *Journal of Business*, 38(1):34–105, 1965.

[6] E. Fama. Random walks in stock market prices. *Financial Analysts Journal*, 21:55–59, 1965.

[7] E. Fama. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25:383–417, 1970.

[8] J. Heer, N. Kong, and M. Agrawala. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *27th International Conference on Human Factors in Computing Systems (CHI)*, pages 1303–1312, 2009.

[9] A. A. Hung, A. Heinberg, and J. K. Yoong. Do risk disclosures affect investment choice? Technical report, RAND Labor and Population, September 2010.

[10] E. J. Keogh and S. Kasetty. On the need for time series data mining benchmarks: A survey and empirical demonstration. *Data Mining and Knowledge Discovery*, 7(4):349–371, 2003.

[11] J. Kozup, E. Howlett, and M. Pagano. The effects of summary information on consumer perceptions of mutual fund characteristics. *Journal of Consumer Affairs*, 42(1):37–59, Spring 2008.

[12] Y. Kroll, H. Levy, and A. Rapoport. Experimental tests of the mean-variance model for portfolio selection. *Organizational Behavior and Human Decision Processes*, 42(3):388–410, 1988.

[13] A. Lo, H. Mamaysky, and J. Wang. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, LV(4):1705–1765, August 2000.

[14] A. W. Lo and A. C. MacKinlay. Stock market prices do not follow random walks: Evidence from a simple specification test. *Review of Financial Studies*, 1(1):41–66, 1988.

[15] A. W. Lo and A. C. MacKinlay. *A Non-Random Walk Down Wall Street*. Princeton University Press, Princeton, NJ, 1999.

[16] B. G. Malkiel. *A Random Walk Down Wall Street*. W. W. Norton & Company, 2003.

[17] H. V. Roberts. Stock-market 'patterns' and financial analysis: Methodological suggestions. *The Journal of Finance*, 14(1):1–10, 1959.

[18] P. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6:41–49, 1965.

[19] L. E. Swedroe. *The Only Guide to a Winning Investment Strategy You'll Ever Need*. St. Martin's Press, 2005.

[20] A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.

[21] K. E. Wärneryd. *Stock-market psychology: How people value and trade stocks*. Edward Elgar, 2001.

[22] H. Wickham, D. Cook, H. Hofmann, and A. Buja. Graphical inference for infovis. *IEEE Trans. Vis. Comput. Graph.*, 16(6):973–979, 2010.

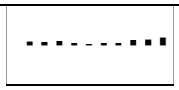
A Appendix

Modified Comparative Chart, Version 2
(Bar Chart of Annual Total Returns)

Performance Information

This table describes investment options that provide variable rates of return. This table shows each option's performance over several time periods and compares the performance with a recognized benchmark. For options with returns that vary over time, past performance does not guarantee how your investment in the option will perform in the future; your investment in these options could lose money.

Table--Variable Return Investments

Name/ Type of Option	Graph: 2000-2009 Year-End Total Returns*	Average Annual Total Return as of 06/30/2010 **1yr.**	**5yr.**	**10yr.**	**Since Inception**	Benchmark/Index as of 06/30/2010 1yr.	5yr.	10yr.
Stock Funds								
Small Cap Stock Index Fund A		22.7%	2.2%	3.7%	5.0% 05/21/98	22.9%	2.2%	3.6% Spliced Small Cap Stock Index
Stock Market Index Fund B		14.3%	-0.9%	-1.7%	10.1% 08/31/76	14.4%	-0.8%	-1.6% S&P 500 Index
Global Equity Fund C		14.5 %	0.2%	4.8%	7.3% 08/14/95	11.8%	1.3%	0.0% Spliced Global Equity Index
Bond Funds								
Bond Market Index Fund D		9.3%	5.5%	6.2%	6.9% 12/11/86	9.5%	5.5%	6.5% Barclays US Aggregate Bond Index
Other								
Money Market Fund E		0.00%	2.7%	2.6%	3.5% 12/14/92	0.0%	2.0%	2.1% Avg. Money Market Treasury Fund
Balanced Index Fund F		13.6%	2.4%	2.4%	7.3% 11/9/92	13.8%	2.4%	2.5% Balanced Composite Index

* The bar charts show the changes in each fund's performance from year to year.

Figure A.1: Picture from Hung, Heinberg, and Yoong (2010).

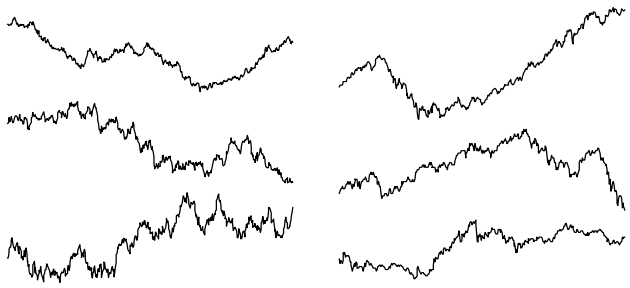


Figure 1. Six time series, three are random walk data, and three are real S&P500 stocks. Experiments suggest that humans cannot tell real and synthetic stock data apart (all the sequences on the right are real).

Table 1. The confusion matrix for human experts in attempting to differentiate between random walk data and stock market data.

		Predicted	
		S&P Stock	**Random Walk**
Actual	**S&P Stock**	20	16
	Random Walk	16	20

Figure A.2: Picture from Keogh and Kasetty (2003).

ARORA is a web game about randomness in financial data.

Four new contests, each worth $25, are now available.

Contests close on December 20th, 2009 at 12pm.

Don't miss your chance to win $100!

This is how the game works:
We collect data from various financial markets and we show it to you in two windows,
- one window plots the real data,
- the other plots the data randomly permuted.

The game asks you to click on the real, non-random data.

To enter a contest, and for your input to be recorded, you must create an account. You will remain anonymous.

Figure A.3: Call for subjects for ARORA.

Ret type	Data type	Mean	Std	1%	25%	50%	75%	99%	Mean	Std	1%	25%	50%	75%	99%
		1st order autocorrelation (%)							*p*-value of Ljung-Box Q_{20} (%)						
		Nasdaq Composite Index (Bear)													
r	**actual**	**25**	10	7	17	23	31	52	**4**	15	0	0	0	1	80
	rand	**0**	5	-10	-2	1	3	11	**51**	33	1	19	57	77	100
r^2	**actual**	**8**	9	-1	2	4	13	39	**51**	46	0	0	66	100	100
	rand	**-1**	3	-5	-2	-1	-1	13	**81**	36	0	90	100	100	100
r^3	**actual**	**5**	8	-4	0	1	7	38	**61**	46	0	0	98	100	100
	rand	**0**	4	-8	-1	0	0	15	**84**	35	0	98	100	100	100
r^4	**actual**	**2**	7	-2	0	0	2	35	**67**	45	0	2	100	100	100
	rand	**0**	2	-2	-1	-1	0	10	**84**	35	0	99	100	100	100
		US Dollar Index (Elk)													
r	**actual**	**-25**	9	-45	-32	-24	-19	-6	**4**	13	0	0	0	1	78
	rand	**-1**	4	-12	-3	-1	2	11	**58**	31	5	27	59	87	99
r^2	**actual**	**15**	9	-1	9	13	19	44	**19**	30	0	0	2	25	100
	rand	**-1**	4	-10	-3	-1	1	15	**63**	30	1	43	70	88	100
r^3	**actual**	**-13**	11	-40	-19	-12	-5	9	**27**	39	0	0	1	60	100
	rand	**0**	4	-9	-2	0	2	13	**71**	35	0	50	88	99	100
r^4	**actual**	**10**	12	-2	2	6	14	56	**48**	43	0	0	46	97	100
	rand	**1**	5	-4	-2	-1	1	24	**72**	37	0	52	93	100	100
		Russell 2000 Index (Bull)													
r	**actual**	**39**	13	10	30	40	48	65	**5**	13	0	0	0	0	66
	rand	**0**	6	-12	-4	0	4	16	**60**	31	0	34	65	90	100
r^2	**actual**	**8**	12	-5	0	5	14	52	**71**	37	0	45	94	100	100
	rand	**-2**	4	-9	-4	-2	0	16	**83**	29	4	81	100	100	100
r^3	**actual**	**6**	10	-2	-1	2	9	51	**80**	35	0	84	100	100	100
	rand	**1**	6	-8	-1	-1	0	27	**87**	28	0	95	100	100	100
r^4	**actual**	**2**	9	-4	-1	-1	0	46	**86**	32	0	99	100	100	100
	rand	**-1**	2	-6	-2	-1	-1	5	**89**	27	0	100	100	100	100
		Gold Spot Price (Reindeer)													
r	**actual**	**-14**	11	-37	-21	-14	-4	13	**15**	24	0	0	2	20	92
	rand	**0**	4	-9	-3	0	2	11	**52**	30	0	24	54	80	99
r^2	**actual**	**10**	8	-2	4	8	15	36	**32**	34	0	0	21	62	100
	rand	**0**	4	-9	-3	0	2	12	**63**	34	0	37	75	94	100
r^3	**actual**	**-7**	9	-42	-11	-5	0	7	**50**	43	0	0	53	97	100
	rand	**-1**	3	-13	-2	0	1	7	**73**	35	0	50	94	100	100
r^4	**actual**	**4**	8	-2	0	1	5	42	**67**	43	0	12	97	100	100
	rand	**0**	3	-4	-2	-1	0	14	**85**	30	0	90	100	100	100

Table A.1: Properties of training daily datasets. For each of the daily datasets and their randomized counterparts used for training, we compute 1st order autocorrelation coefficients, and p-values of the Ljung-Box Q_{20} statistics for various moments of the returns, over nonoverlapping windows of the data corresponding to charts on which subjects were taking a stand. We report mean, standard deviation, and percentiles of these chart-by-chart autocorrelations and p-values (Q_{20}).

Ret type	Data type	Mean	Std	1%	25%	50%	75%	99%	Mean	Std	1%	25%	50%	75%	99%
		1st order autocorrelation (%)							*p*-value of Ljung-Box Q_{20} (%)						
		Dow Jones Industrial Average (Beaver)													
r	**actual**	**1**	7	-6	-5	-1	6	17	**32**	30	0	2	35	49	95
	rand	**0**	5	-10	-4	1	4	8	**52**	29	5	31	50	85	93
r^2	**actual**	**15**	14	-3	6	13	18	45	**12**	27	0	0	0	19	97
	rand	**0**	4	-6	-2	-1	3	8	**58**	35	13	20	54	99	100
r^3	**actual**	**-6**	9	-23	-12	-3	-1	8	**49**	48	0	1	50	95	100
	rand	**-1**	3	-5	-2	-1	0	5	**79**	34	8	38	100	100	100
r^4	**actual**	**10**	17	-2	0	1	14	59	**47**	49	0	0	24	100	100
	rand	**0**	2	-3	-2	-1	0	5	**75**	41	0	33	100	100	100
		Dow Jones Corporate Bond Price Index (Seal)													
r	**actual**	**24**	18	-21	13	23	37	53	**3**	10	0	0	0	0	31
	rand	**-2**	4	-8	-5	-1	1	5	**55**	32	7	29	58	86	95
r^2	**actual**	**29**	14	9	16	26	40	56	**3**	14	0	0	0	0	59
	rand	**-1**	3	-10	-3	-2	-1	6	**85**	22	20	73	95	100	100
r^3	**actual**	**1**	27	-50	-17	14	17	45	**13**	32	0	0	0	3	100
	rand	**-1**	1	-3	-1	0	0	2	**76**	36	0	68	99	100	100
r^4	**actual**	**21**	18	0	5	16	32	56	**28**	43	0	0	0	71	100
	rand	**0**	2	-3	-1	0	1	8	**88**	27	1	89	100	100	100
		Canada/U.S. Foreign Exchange Rate (Lynx)													
r	**actual**	**8**	14	-12	-5	9	17	28	**36**	21	0	23	35	55	64
	rand	**1**	7	-6	-5	-1	7	13	**52**	33	3	29	55	78	100
r^2	**actual**	**21**	18	-1	9	14	29	54	**16**	28	0	0	0	17	85
	rand	**-2**	4	-7	-4	-3	2	5	**80**	23	37	64	89	100	100
r^3	**actual**	**5**	22	-15	-7	-5	10	44	**58**	45	0	3	85	98	100
	rand	**-2**	4	-13	-2	-1	0	2	**69**	44	0	29	99	100	100
r^4	**actual**	**12**	16	-5	1	7	19	44	**67**	40	0	52	81	99	100
	rand	**-1**	3	-5	-2	-2	-1	7	**100**	1	98	99	100	100	100
		S&P GSCI Corn Index Spot (Mandrill)													
r	**actual**	**1**	8	-16	-6	2	7	15	**40**	30	0	14	38	63	99
	rand	**-1**	11	-17	-12	2	8	19	**43**	31	4	17	36	70	99
r^2	**actual**	**10**	12	-12	3	10	15	42	**27**	38	0	0	2	62	99
	rand	**1**	9	-19	-7	2	6	16	**58**	31	0	42	51	87	100
r^3	**actual**	**2**	9	-23	-3	0	9	21	**47**	38	0	10	36	91	100
	rand	**0**	4	-9	-4	0	3	8	**71**	31	8	47	84	99	100
r^4	**actual**	**7**	13	-10	-2	4	13	48	**38**	41	0	1	21	85	100
	rand	**0**	8	-10	-6	-3	5	16	**81**	26	14	76	89	100	100

Table A.2: Properties of training tick datasets. For each of the tick datasets and their randomized counterparts used for training, we compute 1st order autocorrelation coefficients, and p-values of the Ljung-Box Q_{20} statistics for various moments of the returns, over nonoverlapping windows of the data corresponding to charts on which subjects were taking a stand. We report mean, standard deviation, and percentiles of these chart-by-chart autocorrelations and p-values (Q_{20}).