An Experimental Test of Observational Learning Under Imperfect Information*

Boğaçhan Çelen†
Columbia Business School

Shachar Kariv‡
University of California, Berkeley

June 26, 2004.

Abstract

Nearly all observational learning models assume that individuals can observe all the decisions that have previously been made. In reality, such *perfect information* is rarely available. To explore the difference between observational learning under perfect and *imperfect information*, this paper takes an experimental look at a situation in which individuals learn by observing the behavior of their immediate predecessors. Our experimental design uses the procedures of Çelen and Kariv (2004a) and is based on the theory of Çelen and Kariv (2004b). We find that imitation is much less frequent when subjects have imperfect information, even less frequent than the theory predicts. Further, while we find strong evidence that under perfect information a form of generalized Bayesian behavior adequately explains behavior in the laboratory, under imperfect information behavior is not consistent even with this generalization of Bayesian behavior.(*JEL* C92, D8).

*We completed most of the work on this paper when we were both graduate students at New York University. This research was supported by the Center for Experimental Social Sciences (C.E.S.S.) and the C. V. Starr Center for Applied Economics at New York University. We are grateful to Andrew Schotter for his guidance and to an anonymous referees for his comments. We benefited from the expositionary suggestions of William Baumol. We also acknowledge helpful discussions of Colin Camerer, Liran Einav, Xavier Gabaix, Douglas Gale, Charles Holt, David Laibson, and Matthew Rabin. This paper has also benefited from suggestions by the participants of the 2002 International ESA Meeting and seminars at several universities.

†Columbia Business School, 3022 Broadway, Uris Hall, New York, NY 10027 (e-mail: bc2132@columbia.edu).

‡Department of Economics, UC Berkeley, Evans Hall # 3880, Berkeley, CA 94720 (e-mail: kariv@berkeley.edu, url: http://socrates.berkeley.edu/~kariv/).

1 Introduction

Consider a sequence of individuals who make a once-in-a-lifetime decision under incomplete and asymmetric information. If each decision is announced publicly, and thus is known to all successors, despite the asymmetry of information, eventually individuals will imitate their predecessor's behavior even if it conflicts with their private information. In other words, individuals 'ignore' their own information and follow a herd. Furthermore, since actions aggregate information poorly, the prediction of the theory, which was matched in many experiments, is that herds are likely to form and adopt a suboptimal action. These are the main results of the observational learning literature introduced by Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992).[1]

A central assumption of nearly all observational learning models is *perfect information*: everyone is assumed to be informed about the entire history of actions that have already been taken. In reality, individuals have imperfect information. If each individual observes the actions of only a small number of other individuals, it is not clear that herd behavior will arise. In Çelen and Kariv (2004b), we abandon the perfect-information assumption and explore behavior when each individual observes only her immediate predecessor's decision.

Our imperfect-information model provides outcomes that are quite distinct from and in some ways more extreme than the perfect-information model. We predict longer and longer periods of uniform behavior, punctuated by (increasingly rare) switches. Thus, the perfect- and imperfect-information versions of the model share the conclusion that individuals can, for a long time, make the same choice. The important difference is that, whereas in the perfect-information model a herd is an absorbing state, in the imperfect-information model, there are continued, occasional and sharp shifts in behavior.

It is the goal of this paper to explore behavior under imperfect information experimentally and to provide a comparison with the results obtained under perfect information by Çelen and Kariv (2004a). To this end, we use an experimental design, that allows an environment richer and more flexible

[1] See Chamley (2003). For excellent surveys see, Gale (1996) and Bikhchandani, Hirshleifer and Welch (1998), which also provide examples and applications of observational learning in economic contexts. There are further extensions to the theory, notably, Lee (1993), Chamley and Gale (1994), Gul and Lundholm (1995), and Smith and Sørensen (2000).

than the one in the existing literature.[2]

In the experiment, a sequence of subjects draw private signals from a uniform distribution over $[-10, 10]$. The decision problem is to predict whether the sum of all subjects' signals is positive or negative and to choose an appropriate action, A or B. A is the profitable action when this sum is positive and B if it is not. However, instead of choosing action A or B directly, after being informed about the decision of the preceding subject and before observing their own private signals, subjects are asked to select a cutoff such that action A will be chosen if the signal received is greater than the cutoff and action B will be selected otherwise. Only after a subject reports her cutoff, is she informed of her private signal, and her action is recorded accordingly.

Aside from the information structure, this experimental design is identical to the one employed in Çelen and Kariv (2004a). That is, both experiments employ the same procedures but the histories of actions observed by subjects are different, in fact, at the opposite extreme. For comparison purposes, our new results will be presented along with the results of Çelen and Kariv (2004a). Thus, this paper offers two contributions to methodology: First, it shows how to deal with the case in which each subject can observe only her immediate predecessor's decision, an information structure hitherto unexplored in experimental studies. Second, it uses a cutoff elicitation technique such that, instead of making a decision *per se*, subjects state a cutoff that then determines their action.

We find that imitation is much less frequent when subjects have imperfect information, even less frequent than the theory predicts. For a better understanding of the decision mechanism of the subjects, we focus on the data at the individual level. We find that among the subjects who follow their predecessor there is a good degree of conformity with the theory, which we fail to observe in the aggregate data. Under imperfect information decision-making is of course much more complex, and thus, mistakes are more likely to occur.

For that reason we turn to the robustness of the existing theory by tackling its central assumption of common knowledge of rationality. We introduce a model that explains subjects' behavior as a form of generalized Bayesian behavior that incorporates limits on the rationality of others. While we find strong evidence that this form of generalized Bayesian behav-

[2] Anderson and Holt (1997) investigate the model of Bikhchandani, Hirshleifer and Welch (1992) experimentally. Following their pioneering work, Allsopp and Hey (2000), Anderson (2001), Hung and Plott (2001), and Kübler and Weizsäcker (2004), among others, analyze observational learning under perfect information.

ior adequately explains behavior in the laboratory under perfect information, under imperfect information behavior is not consistent even with this generalization.

Finally, we discuss an alternative theoretical explanation for the difference in the results, by introducing noisy individuals in the benchmark model. We find that under perfect information noise does not jeopardize the learning of the rational individuals, whereas under imperfect information learning breaks down since the information revealed by the actions of rational individuals is lost to the background noise.

The paper is organized as follows. The next section describes the experimental design and procedures, and section 3 outlines the underlying theory. Section 4 summarizes the results and provides an econometric analysis. Section 5 discusses the results. Section 6 contains some concluding remarks.

2 Experimental Design

The procedures described below are identical to those used by Çelen and Kariv (2004a) with the exception that the history of actions observed by subjects is different. The experiment was run at the Experimental Economics Laboratory of the Center for Experimental Social Sciences (C.E.S.S.) at New York University. The 40 subjects in this experiment were recruited from undergraduate economics classes at New York University and had no previous experience in observational learning experiments. In each session eight subjects participated as decision-makers. After subjects read the instructions (the instructions are available upon request) they were also read aloud by an experimental administrator. The experiment lasted for about one and a half hours. A $5 participation fee and subsequent earnings for correct decisions, which averaged about $19, were paid in private at the end of the session. Throughout the experiment we ensured anonymity and effective isolation of subjects in order to minimize any interpersonal influences that could stimulate uniformity of behavior.

Each experimental session entailed 15 independent rounds, each divided into eight decision-turns. In each round, all eight subjects took decisions sequentially in a random order. A round started by having the computer draw eight numbers from a uniform distribution over $[-10, 10]$. The numbers drawn in each round were independent of each other and of the numbers in any of the other rounds. Each subject was informed only of the number corresponding to her turn to move. The value of this number was her private signal. In practice, subjects observed their signals up to two decimal points.

Upon being called to participate, a subject first observed the action taken by the preceding subject in that round. After this and before being informed of her private signal, each subject was asked to select a number between -10 and 10 (a cutoff), for which she would take action A if her signal was above the cutoff and action B if it was not. Action A was profitable if and only if the sum of the eight numbers was positive and action B was profitable otherwise. Only after submitting her cutoff, the computer informed her of the value of her private signal. Then, the computer recorded her decision as A if the signal was higher than the cutoff she selected. Otherwise, the computer recorded her action as B.

After all subjects had made their decisions, the computer informed everyone what the sum of the eight numbers actually was. All participants whose decisions determined A as their action earned \$2 if this sum was positive (or zero) and nothing otherwise. Similarly, all whose decisions led to action B earned \$2 if this sum was negative and nothing otherwise. This process was repeated in all rounds. Each session was terminated after all 15 rounds were completed.

3 Some Theory

3.1 The Bayesian solution

In this section we discuss at some length the theoretical predictions of the model tested in the laboratory. Çelen and Kariv (2004b) provides an extensive analysis of a more general version of the model.

To formulate the Bayesian solution of the decision problem underlying our experimental design, suppose that the eight individuals receive private signals $\theta_1, \theta_2, .., \theta_8$ that are independently and uniformly distributed on $[-1, 1]$.[3] Sequentially, each individual $n \in \{1, ..., 8\}$ has to make a binary irreversible decision $x_n \in \{A, B\}$ where action A is profitable if and only if $\sum_{i=1}^{8} \theta_i \geq 0$, and action B otherwise. Furthermore, except the first individual, everyone observes only her immediate predecessor's decision.

In such a situation, conditional on the information available to her, individual n's optimal decision rule is

$$x_n = A \text{ if and only if } \mathbb{E}\left[\sum\nolimits_{i=1}^{8} \theta_i \mid \theta_n, x_{n-1}\right] \geq 0$$

[3] For expository ease, we normalize the signal space to $[-1, 1]$.

and since individuals do not know any of their successors' actions,

$$x_n = A \text{ if and only if } \theta_n \geq -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid x_{n-1}\right].$$

It readily follows that the optimal decision takes the form of the following *cutoff strategy*,

$$x_n = \begin{cases} A & \text{if} \quad \theta_n \geq \hat{\theta}_n, \\ B & \text{if} \quad \theta_n < \hat{\theta}_n, \end{cases} \tag{1}$$

where

$$\hat{\theta}_n = -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid x_{n-1}\right] \tag{2}$$

is the optimal cutoff which accumulates all the information revealed to individual n from her predecessor's action. Since $\hat{\theta}_n$ is sufficient to characterize individual n's behavior, the sequence of cutoffs $\{\hat{\theta}_n\}$ characterizes the social behavior. We take these as the primitives of the experimental design and of our analysis.

We proceed by illustrating the basic features of the decision problem. The first individual's decision is based solely on her private signal. Therefore, her optimal cutoff is $\hat{\theta}_1 = 0$ meaning that it is optimal for her to take action A if and only if $\theta_1 \geq 0$ and action B otherwise. Since the second individual observes the first's action, she conditions her decision on whether $x_1 = A$ or $x_1 = B$. If, for example, $x_1 = A$, then $\mathbb{E}[\theta_1 | x_1 = A] = 1/2$ and thus it is optimal for the second individual to take action A if and only if $\theta_2 \geq -1/2$. Likewise, if $x_1 = B$ it is optimal for her to take action A if and only if $\theta_2 \geq 1/2$. Thus, according to (2) the second individual's cutoff rule is

$$\hat{\theta}_2 = \begin{cases} -\frac{1}{2} & \text{if} \quad x_1 = A, \\ \frac{1}{2} & \text{if} \quad x_1 = B. \end{cases} \tag{3}$$

Note that for any $\theta_2 \in [-1/2, 1/2)$ the second individual imitates the first even though she would have taken a contrary action had she based her decision solely on her own signal.

By the time it is the third individual's turn to make a decision, the information inherent in the first individual's action is suppressed, but she can still draw a probabilistic conclusion about it by Bayes' rule. That is, by observing the action of the second individual the third assigns probability to the actions that the first individual could have taken. For example, by observing $x_2 = A$, she assigns probability $3/4$ that $x_1 = A$ and probability $1/4$ that $x_1 = B$. A simple computation shows that $\mathbb{E}[\theta_1 + \theta_2 | x_2 = A] = 5/8$ which implies that if $x_2 = A$ it is optimal for the third individual to take action A for any signal $\theta_3 \geq -5/8$. A similar analysis shows that if $x_2 = B$ it is optimal for her to take action A for any signal $\theta_3 \geq 5/8$. Thus, according to (2) the third individual's cutoff rule is

$$\hat{\theta}_3 = \begin{cases} -\frac{5}{8} & \text{if} \quad x_2 = A, \\ \frac{5}{8} & \text{if} \quad x_2 = B. \end{cases} \tag{4}$$

Note that the action of the second individual reflects part of the information of the first individual, so relative to the first individual's action more information is revealed by the second's action. For that reason, the third individual is *ex ante* more likely to act like her predecessor than the second individual. For example, if the first individual takes action A, then by (3) the second individual imitates her for any private signal $\theta_2 \in [-1/2, 1]$. Whereas, if the second individual takes action A, according to (4), the third individual imitates the second for any private signal $\theta_3 \in [-5/8, 1]$.

Proceeding with the example by adding individuals who receive private signals and learn only from preceding individual's action, the cutoff rule, $\hat{\theta}_n$, of any individual n can take the two different values conditional on whether individual $n-1$ took action A or action B which we denote by

$$\begin{aligned} \overline{\theta}_n &= -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid x_{n-1} = A\right], \\ \underline{\theta}_n &= -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid x_{n-1} = B\right]. \end{aligned}$$

Note that if individual n observes $x_{n-1} = A$, she can determine the probabilities that $x_{n-2} = A$ or $x_{n-2} = B$ conditional on this information. If $x_{n-2} = 1$ then the actual cutoff of individual $n-1$ is $\overline{\theta}_{n-1}$, which already inherits all the information accumulated in the history. Moreover, the expected value of her signal θ_{n-1} is computable conditional on $\overline{\theta}_{n-1}$ and $x_{n-1} = A$. Using these observations, In Çelen and Kariv (2004b) we show that the law of motion for $\overline{\theta}_n$ is given by

$$\begin{aligned} \overline{\theta}_n = P(x_{n-2} = A | x_{n-1} = A) \left\{\overline{\theta}_{n-1} - \mathbb{E}\left[\theta_{n-1} \mid x_{n-2} = B\right]\right\} \\ + P(x_{n-2} = B | x_{n-1} = B) \left\{\underline{\theta}_{n-1} - \mathbb{E}\left[\theta_{n-1} \mid x_{n-2} = B\right]\right\}, \end{aligned}$$

which simplifies to

$$\overline{\theta}_n = \frac{1 - \overline{\theta}_{n-1}}{2}\left[\overline{\theta}_{n-1} - \frac{1 + \overline{\theta}_{n-1}}{2}\right] + \frac{1 - \underline{\theta}_{n-1}}{2}\left[\underline{\theta}_{n-1} - \frac{1 + \underline{\theta}_{n-1}}{2}\right]. \tag{5}$$

An analogous argument also applies for the law of motion for $\underline{\theta}_n$. Using symmetry, $\overline{\theta}_n = -\underline{\theta}_n$, the dynamics of the cutoff rule $\hat{\theta}_n$ is described in a closed form solution recursively as follows:

$$\hat{\theta}_n = \begin{cases} -\frac{1+\hat{\theta}_{n-1}^2}{2} & \text{if} \quad x_{n-1} = A, \\ \frac{1+\hat{\theta}_{n-1}^2}{2} & \text{if} \quad x_{n-1} = B, \end{cases} \tag{6}$$

where $\hat{\theta}_1 = 0$.

The impossibility of an informational cascade follows immediately from (6) since for every n, $-1 < \hat{\theta}_n < 1$. That is, in making a decision, everyone takes her private signal into account in a non-trivial way. However, as Figure 1 illustrates, according to (6) the cutoff rule partitions the signal space into three subsets: $[-1, \overline{\theta}_n), [\overline{\theta}_n, \underline{\theta}_n)$ and $[\underline{\theta}_n, 1]$. For high-value signals $\theta_n \in [\underline{\theta}_n, 1]$ and low-value signals $\theta_n \in [-1, \overline{\theta}_n)$ individual n follows her private signal and takes action A or B respectively. In the intermediate subset $[\overline{\theta}_n, \underline{\theta}_n)$, which we call an *imitation set*, private signals are 'ignored' in making a decision and individuals imitate their immediate predecessor's action. Furthermore, since $\{\overline{\theta}_n\}$ and $\{\underline{\theta}_n\}$ converge respectively to -1 and 1, imitation sets monotonically increase in n regardless of the actual history of actions, and thus, over time, it is more likely that imitation will arise.

[Figure 1 here]

In fact, in Çelen and Kariv (2004b) we show that when the population is arbitrary large the imitation sets converge to the entire signal space in the limit. However, note that this does not imply convergence of the cutoff process (6). A careful analysis shows that it is not stable either at -1 or at 1. This implies that there will always be an individual who will choose an action different from her predecessor's because of a contrary signal. Therefore, herd behavior is impossible. However, although there is no convergence of actions in the standard herding manner, the behavior exhibits longer and longer periods in which individuals act alike, punctuated by increasingly rare switches.

3.2 A note on perfect and imperfect information

Next, we shall investigate the differences between the decision problem under perfect and imperfect information. Under perfect information, the optimal decision also takes the form of the cutoff strategy given in (1) where the cutoff rule is a function of the entire realized history of actions:

$$\hat{\theta}_n = -\mathbb{E}\left[\sum\nolimits_{i=1}^{n-1} \theta_i \mid (x_i)_{i=1}^{n-1}\right]. \tag{7}$$

Since under perfect information any history is shared as public information, individual n's cutoff $\hat{\theta}_n$ can be inferred perfectly by her successors. In other words, everyone can deduce what each of her predecessors has learned. As a result, under perfect information the cutoff rule exhibits the following recursive structure,

$$\hat{\theta}_n = \hat{\theta}_{n-1} - \mathbb{E}[\theta_{n-1} \mid \hat{\theta}_{n-1}, x_{n-1}],$$

which results in the following cutoff process, described recursively.

$$\hat{\theta}_n = \begin{cases} \frac{-1+\hat{\theta}_{n-1}}{2} & \text{if} \quad x_{n-1} = A, \\ \frac{1+\hat{\theta}_{n-1}}{2} & \text{if} \quad x_{n-1} = B, \end{cases} \tag{8}$$

where $\hat{\theta}_1 = 0$.

As in the imperfect information case, the impossibility of an informational cascade follows immediately, since for any individual n $-1 < \hat{\theta}_n < 1$. However, the cutoff process has the martingale property $\mathbb{E}[\hat{\theta}_{n+1} \mid \hat{\theta}_n] = \hat{\theta}_n$, so by the Martingale Convergence Theorem it is stochastically stable in the neighborhood of the fixed points, -1 and 1. Further, since convergence of the cutoff process implies convergence of actions, behavior settles down in finite time. Hence, under perfect information, a cascade cannot arise but herd behavior must.[4]

Let us fix ideas in terms of the preceding illustration. Under perfect information, since the first individual's action is public information known to both successors, the third individual knows the observation on which the second based her decision. Thus, according to (8), a simple computation yields that the third's cutoff rule is given by

$$\hat{\theta}_3 = \begin{cases} -\frac{3}{4} & \text{if} \quad x_1 = A, x_2 = A, \\ -\frac{1}{4} & \text{if} \quad x_1 = B, x_2 = A, \\ \frac{1}{4} & \text{if} \quad x_1 = A, x_2 = B, \\ \frac{3}{4} & \text{if} \quad x_1 = B, x_2 = B. \end{cases}$$

If we add individuals and proceed with the same analysis, we find that if the first three individuals choose A, the fourth individual's cutoff is $\hat{\theta}_4 = -7/8$; if the first four individuals choose A, the fifth individual's cutoff is $\hat{\theta}_5 = -15/16$; and so on. Hence, any successive individual who also chooses

[4] An informational cascade is said to occur when an infinite sequence of individuals ignore their private information when making a decision, whereas herd behavior occurs when an infinite sequence of individuals make an identical decision, not necessarily ignoring their private information.

action A reveals less of her private information and makes it more difficult for her predecessor not to choose action A.

On the other hand, if the fourth individual chooses action B after the first three individuals choose A, her decision reveals that her signal lies in the interval $[-1, -7/8)$ and the fifth individual's cutoff is $\hat{\theta}_5 = 1/16$. Hence, the longer a cluster of individuals acts alike, the larger the asymmetry between the information revealed by imitation and deviation. Notice that a deviator induces her successor to be slightly in favor of joining the deviation, which is referred in the literature as the *overturning principle*.

Under imperfect information, in contrast, the overturning principal has a more extreme nature. To illustrate, suppose that the first three individuals take action A. Thus, according to (6) the fourth individual's cutoff is $\hat{\theta}_4 = -0.695$. Now, if the fourth individual has a contrary signal, $\theta_4 \in (-0.695, -1]$, she deviates by choosing action B. Moreover, since the deviation is not observed by the fifth individual, she sharply overturns behavior by setting her cutoff near 1, specifically at $\hat{\theta}_5 = 0.743$. Hence, deviation of the fourth individual makes it hard for the fifth individual not to follow the deviation.

In conclusion, according to the overturning principle, under both perfect and imperfect information a deviator becomes a leader to her successors. Nevertheless, there is substantial difference. Under perfect information the deviator can be identified since previous actions are publicly known. As a result, her deviation reveals clear cut information regarding her private signal that meagerly dominates the accumulated public information. Thus, her successor will slightly favor joining the deviation. On the other hand, under imperfect information, one cannot tell whether her predecessor is an imitator or a deviator. Thus, the action of the deviator is her successor's only statistic from which to infer the entire history of actions. Consequently, one who follows a deviator is very enthusiastic to join the deviation.

4 Experimental Results

4.1 Descriptive statistics

4.1.1 Group behavior

We identify a subject who engages in cascade behavior as one who reports a cutoff of -10 or 10, and thus takes either action A or B, no matter what private signal she receives. In contrast, a subject who joins a herd but does not engage in cascade behavior is one whose cutoff is in the interval

$(-10, 10)$, indicating that there are some signals that can lead her to choose action A and some that lead to B but when her private signal is realized she acts as her predecessors did. Finally, we say that a cascade occurs in the laboratory when beginning with some subject, all others thereafter follow cascade behavior, and herd behavior occurs when, beginning with some subject, all take the same action.

Through all of the experimental sessions, we observed herds of at least five subjects in 8 of the 75 rounds (10.7 percent). As Table 1 shows, of these 8 rounds, in 2 rounds all eight subjects acted alike, in 1 round the last six subjects and in 5 rounds the last five subjects acted alike. All herds, except one, were consistent with the optimal cutoff rules given by (2). Moreover, even though subjects had imperfect information about the history of decisions, all herds selected the correct action (defined relative to the information available to the group). The theoretical prediction, in contrast, is that even under imperfect information herds should arise in more than half of the rounds (63.4 percent), yet 19.8 percent of these herds should entail incorrect decisions.[5] Finally, since herds developed rarely, it is clear that overturns in behavior occurred often. Excluding the first decision turn, such overturns happened in 234 of the 525 decisions points (39.0 percent), whereas the theory predicts behavior overturns at only 19.0 percent of the decision points.

[Table 1 here]

Table 1 illustrates the instances where a herd is not the result of an informational cascade. For example, in rounds 1.7 (the seventh round in the first session) and 4.11, an informational cascade did not occur, yet all subjects followed a herd. Although theory predicts that cascades do not occur, we observe them in the laboratory. Informational cascades were observed in 18 rounds (24.0 percent) of which in two rounds the last two subjects followed cascade behavior, and in 16 rounds only the last subject followed cascade behavior. Table 2 summarizes the data and the Bayesian outcome for selected rounds in which cascades occur. In addition, a cascade behavior,

[5] We compute the probability with the help of simulations since, conditional on the state of the world $\sum_{n=1}^{8} \theta_n$, private signals are perfectly negatively correlated, which makes the problem very hard to solve analytically.

The simulations were carried out by MatLab. An experiment starts by drawing a vector of ten *i.i.d.* signals from uniform distribution over $[-10, 10]$. Then, we collect the actions generated by this vector according to cutoff processes. Experiments are repeated until the marginal change in the average number of correct actions for additional 10^7 experiments is less than 10^{-5}.

which was not part of an informational cascade, was observed in 85 decision turns. In total, cascade behavior was observed in 105 out of 600 turns (17.5 percent). However, 65 of these 105 (61.9 percent) entail to a small number of subjects who consistently followed cascade behavior in most rounds in which they participated.[6]

[Table 2 here]

Table 3 summarizes our experimental results and compares them with the results we reported in Çelen and Kariv (2004a). Under perfect information, herds were observed in 27 of the 75 rounds (36.0 percent), and in half of the herds all subjects acted alike. Moreover, all herds except one turned out to be on the correct decision. Perhaps the most unexpected result under perfect information, at least from a theoretical perspective, is that informational cascades were observed in 26 rounds (34.7 percent). Accordingly, we conclude that although from a theoretical point of view cascade behavior is a mistake, it is a behavioral phenomenon. Under imperfect information, in contrast, both herds and cascades are much less frequent. Finally, over all subjects, earnings for correct decisions averaged \$18.8 under imperfect information and \$22.0 under perfect information, a difference of 17.0 percent. A binary Wilcoxon test indicates that there is a significant difference between the sample of subject payoffs under perfect and imperfect information at the 5 percent significance level.

[Table 3 here]

The decrease in the payoffs under imperfect information, relative to those under perfect information, is mainly attributable to the decreasing number of herds. Note that the number of herds observed under imperfect information is 71.4 percent less than the number of herds observed under imperfect information. Remarkably, under both perfect and imperfect information all herds except one turned out to be on the correct decision. This is particularly interesting since the prediction of the theory, which was replicated in many experiments, is that uniform behavior is likely to be erroneous. In Çelen and Kariv (2004a), we argue that possible reasons for the difference is the richness of the continuous signal space, and that subjects can fine-tune their decisions by choosing a cutoff strategy instead of taking an action directly. Simulations, however, suggest that theoretically the probabilities

[6] Of all 40 subjects, two followed a cascade behavior in all rounds, one in 11 rounds, one in nine rounds, one in eight rounds and one in seven rounds.

of ending up in a correct (incorrect) herd are 62.9 percent (20.0 percent) and 50.8 percent (12.5 percent) under perfect and imperfect information respectively.

4.1.2 Individual behavior

To organize our cutoff data and to put them into perspective, we first define decisions made by subjects as *concurring decisions* if the sign of their cutoff agrees with the action taken by their predecessor. For instance, when a subject observes that her predecessor took action A (B) and adopts a negative (positive) cutoff, she demonstrates concurrence, since by selecting a negative (positive) cutoff she adopts a higher probability of taking action A (B). Similarly, if a subject observes action A (B) and selects a positive (negative) cutoff, then she disagrees with her predecessor. We say that such decisions are *contrary decisions.* Finally, *neutral decisions* are carried out by choosing a zero cutoff, which neither agrees nor disagrees with the predecessor's action but simply entails choice based on private information.

Over all decision turns, excluding the first, 44.2, 39.2 and 16.6 percents of the decisions were concurring, contrary and neutral in that order. Thus, subjects tended to follow the actions of their predecessor far less than the theory predicts. In addition to presenting the data on the number of decision points that were concurring, neutral or contrary, we look at the distribution of subjects in terms of the frequency with which they either agreed or disagreed with their predecessor's action. Figure 2 summarizes the percent of subjects who disagreed with the observed action in less than two rounds, three to five rounds and so on. Notice that subjects tended to disagree very often. In fact, only 20.0 percent of the subjects disagreed less than two times and 40.0 percent of the subjects disagreed with the action they observed about half of the times. This is a strong indication that subjects acted in a manner that is not consistent with the prediction of the theory.

[Figure 2 here]

The signs of the cutoffs as indicating agreement or disagreement tells only part of the story as it ignores the strength of this agreement or disagreement, which can be measured by the magnitude of the cutoff set. For example, if one observes action A and sets a cutoff close to -10, then not only she agrees with the action she observed, but she does so very strongly since she will almost surely take action A. In contrast, selection of a negative cutoff that is closer to zero clearly indicates a much weaker agreement.

Since the cutoff strategy is symmetric around zero, we proceed by transforming the data generated by our subjects in the following way: Take the absolute value of cutoffs in concurring decision points and minus the absolute value of cutoffs at contrary decision points. For instance, if a subject observes action A and selects a cutoff of -5, we take it as 5, since she acts in a concurring manner. On the other hand, if she places a cutoff of 5 we take it as -5, since she acts in a contrary manner.

Figure 3 presents the theoretical cutoffs and the mean cutoff of concurring decisions turn by turn. Note that there is a substantial degree of conformity with the theory in the magnitude of the cutoffs chosen by subjects when they agreed with the action observed. In other words, once a subject has decided to imitate her predecessor's action, she does so with the right intensity in the Bayesian sense as the cutoffs chosen are quite close to those the theory predicts. However, Figure 3 shows clearly that the situation reverses, particularly in late decision-turns, when we include neutral decisions in our sample.

[Figure 3 here]

So far, we focused on concurring decisions. There is, however, the complement subset of contrary decisions. Notice that once a subject decides not to follow her predecessor's action, the intensity of her disagreement can be measured in several ways. Figure 4 presents the intensity of disagreement in two ways. First, we use the absolute value of the distance between the cutoff actually chosen and the one which would be selected if the subject acted according to the theoretical cutoff rule, and, second, by the absolute value of the distance of the chosen cutoff from zero. Note that the strength of disagreement is rather severe since when subjects disagree with their predecessor they tend to do so in quite an extreme way.

[Figure 4 here]

All of the results presented above condition our data on whether decisions are concurring or contrary. Figure 5 shows that if we do not condition the data on agreement and disagreement, it appears that overall there is a significant difference from what the theory predicts. In fact, the heuristic in which subjects follow their own signal outperforms Bayesian behavior as a predictor to the behavior in the laboratory. However, the difference from the prediction of the theory is in fact a compositional difference representing the distribution of decisions over our concurring and contrary categories and not differences in how persuasive predecessors' actions are once they are followed.

[Figure 5 here]

The regression analysis presented in Table 4 summarizes our discussion so far. We regress the transformed cutoff set by subjects on the decision turn as well as dummy variables which take a value of one in the first and last five rounds in a session.[7] Note that cutoffs are not expected to increase with later turns as every coefficient is not significantly different from zero. Thus, the regression clearly indicates that, subjects, when they repeat the rounds, are not increasingly persuaded by the observed action.

[Table 4 here]

Comparing the individual behavior with that reported in Çelen and Kariv (2004a) indicates that perfect information appears to be rationality enhancing. To demonstrate this, under each information structure, for each subject, we compute the mean squared deviation (MSD) between the cutoff a subject reports and that prescribed by the theory. The smaller the mean MSD for subjects in any information structure the closer is their behavior to that predicted by the theory. The histograms in Figure 6 show that subject behavior is more consistent with the theory under perfect information as the distribution of MSD scores shifts considerably to the left when calculated using the perfect information data. A Kolmogorov-Smirnov test confirms this observation at the 5 percent significance level.

[Figure 6 here]

4.1.3 An econometric analysis

In Çelen and Kariv (2004a), in order to explain the behavior in the laboratory, we test a model that describes subjects' behavior as a form of generalized Bayesian behavior that incorporates limits on the rationality of others. We find strong evidence that this type of Bayes rationality explains the behavior in the laboratory. For comparison purposes, we repeat the same exercise here.

We assume that subjects estimate the errors of others and consider this in processing the information revealed by their predecessors' actions. We attempt to formulate this by estimating a recursive model that allows for the possibility of errors in earlier decisions. This approach enables us to

[7] There is no control for subjects' private signals because each subject was asked to select a cutoff after observing the action taken by the preceding subject but before being informed of her private signal.

evaluate the degree to which Bayes rationality explains behavior in the laboratory. Anderson and Holt (1997) also employ this approach, but while they use subjects' expected payoffs, our cutoff elicitation allows us to estimate recursively the process of cutoff determination adjusted for decision errors and independent shocks.

For this purpose, suppose that at each decision turn n, with probability p_n an individual is Bayesian and rationally computes her cutoff, and with probability $(1 - p_n)$, she is noisy, in the sense that her cutoff is a random draw from a distribution function G_n with support $[-1, 1]$ (for expository ease, we again normalize the signal space) and mean $\tilde{\theta}_n$. Suppose that others cannot observe whether an individual's behavior is noisy, but the sequences $\{p_n\}$ and $\{G_n\}$ are common knowledge among individuals. In addition, we assume that rational individuals could tremble, in the sense that their cutoff embodying uncorrelated small computation or reporting mistakes. To be precise, a rational individual in turn n reports cutoff $\hat{\theta}_n + \phi_n$ where ϕ_n is distributed normally with mean 0 and variance σ_n^2. Note that the mistakes of the rational individuals are a tremble from the rational cutoff, i.e., has mean $\hat{\theta}_n$, whereas noisy individuals make decisions randomly.

After adding noisy individuals to the model, the law of motion for $\overline{\theta}_n$ becomes

$$\bar{\theta}_n = -\{p_{n-1}\mathbb{E}[\textstyle\sum_{i=1}^{n-1} \theta_i \mid x_{n-1} = A] + (1 - p_{n-1})\mathbb{E}[\theta_{n-1} \mid G_{n-1}, x_{n-1} = A]\}, \tag{9}$$

where

$$\mathbb{E}[\theta_{n-1} \mid G_{n-1}, x_{n-1} = A] = \int_{-1}^{1} \frac{1+x}{2} dG_{n-1}(x) = \frac{1 + \tilde{\theta}_{n-1}}{2},$$

and by using (5)

$$\begin{aligned}\overline{\theta}_n = p_{n-1}\{&\frac{1-\overline{\theta}_{n-1}}{2}[\overline{\theta}_{n-1} - \frac{1+\overline{\theta}_{n-1}}{2}] \\ &+ \frac{1-\underline{\theta}_{n-1}}{2}[\underline{\theta}_{n-1} - \frac{1+\underline{\theta}_{n-1}}{2}]\} \\ &- (1-p_{n-1})\frac{1+\tilde{\theta}_n}{2}.\end{aligned}$$

An analogous analysis applies for the law of motion for $\underline{\theta}_n$.[8]

[8] Note that we do not assume that G_n satisfies symmetry, i.e., $G_n(\theta) = 1 - G_n(-\theta)$ for any n and $\theta \in [-1, 1]$, so after adding noise generically $\bar{\theta}_n \neq -\underline{\theta}_n$.

Under these assumptions, at any decision turn n and round i, the expected cutoff is

$$y_n^i = (1-p_n)\tilde{\theta}_n + p_n\hat{\theta}_n^i + p_n\phi_n^i,$$

and in matrix form

$$\mathbf{y_n} = (1-p_n)\tilde{\theta}_n\mathbf{1} + p_n\hat{\boldsymbol{\theta}}_\mathbf{n} + p_n\boldsymbol{\phi}_\mathbf{n}, \tag{10}$$

where $\mathbf{y_n}$, $\mathbf{1}$, $\hat{\boldsymbol{\theta}}_\mathbf{n}$ and $\boldsymbol{\phi}_\mathbf{n}$ are vectors whose components are y_n^i, 1, $\hat{\theta}_n^i$ and ϕ_n^i respectively. This leads the following econometric specification:

$$\mathbf{y_n} = \alpha_n\mathbf{1} + \beta_n\mathbf{z_n} + \boldsymbol{\varepsilon}_\mathbf{n}, \tag{11}$$

where

$$\alpha_n = (1-p_n)\tilde{\theta}_{n-1},\ \beta_n = p_n \text{ and } \boldsymbol{\varepsilon}_\mathbf{n} = p_n\boldsymbol{\phi}_\mathbf{n}.$$

For any round i, $\mathbf{z_1} = \mathbf{0}$ and for any turn $n > 1$, the i^{th} component of the vector $\mathbf{z_n}$ is

$$z_n^i = \begin{cases} \overline{z}_n & \text{if} \quad x_{n-1}^i = A, \\ \underline{z}_n & \text{if} \quad x_{n-1}^i = B, \end{cases} \tag{12}$$

where

$$\begin{aligned} \overline{z}_n^i = \hat{\beta}_{n-1}\{&\frac{1-\overline{z}_{n-1}^i}{2}[\overline{z}_{n-1}^i - \frac{1+\overline{z}_{n-1}^i}{2}] \\ &+ \frac{1-\underline{z}_{n-1}^i}{2}[\underline{z}_{n-1}^i - \frac{1+\underline{z}_{n-1}^i}{2}]\} \\ &- \frac{1-\hat{\beta}_{n-1}+\hat{\alpha}_{n-1}}{2}. \end{aligned}$$

A similar analysis also applies for $\underline{z}_n$.[9]

Notice that the parameters are estimated recursively. That is, the estimated parameters for the first decision-turn, $\hat{\alpha}_1$ and $\hat{\beta}_1$, are employed in estimating the parameters for the second turn, α_2 and β_2, and so on. So, at each turn n, the estimates for the previous turn $\hat{\alpha}_{n-1}$ and $\hat{\beta}_{n-1}$ are used

[9] A similar econometric specification (11) is employed in Çelen and Kariv (2004a) under perfect information, but for any turn $n > 1$ the error-adjustment updating rule (12) exhibits the following recursive structure

$$z_n^i = z_{n-1}^i - \begin{cases} \frac{1+(\hat{\alpha}_{n-1}+\hat{\beta}_{n-1}z_{n-1}^i)}{2} & \text{if} \quad x_{n-1}^i = A, \\ \frac{-1+(\hat{\alpha}_{n-1}+\hat{\beta}_{n-1}z_{n-1}^i)}{2} & \text{if} \quad x_{n-1}^i = B. \end{cases}$$

to calculate an estimate of the optimal cutoff for each decision $\bar{\theta}_n^i$ or $\underline{\theta}_n^i$, denoted respectively by $\bar{z}_n^i$ and $\underline{z}_n^i$, which, in turn, constitutes the independent variable in the estimation (11) for that turn.

Coefficient β is the probability that a subject participating in decision-turn n is rational, which can be interpreted as a parameterization of the average weights given to the information revealed by the history of actions. On the other hand, coefficient α can be interpreted as a parameterization of the information processing bias such as a blind tendency toward a particular action. For example, since $\tilde{\theta}_n = \alpha_n/(1-\beta_n)$, when $\beta_n < 1$, any $\alpha_n < 0$ ($\alpha_n > 0$) indicates that subjects participating in turn n are biased toward action A (B).

When the information processing biases diminish, i.e., $\alpha_n \to 0$, and $\beta_n \to 1$ (and $\sigma_n^2 \to 0$), the behavior tends to become Bayesian. That is, when $\alpha_n = 0$ and $\beta_n = 1$ for all n, according to (11), the laboratory decision-making conforms perfectly with the optimal history-contingent cutoff process given by (6). Similarly, the behavior tends to be random as $\alpha_n \to 0$ and $\beta_n \to 0$. Notice that when $\alpha_n = \beta_n = 0$ (and $\sigma_n^2 \to 0$), equation (11) requires expected cutoff to be zero, which is simply a choice based on private information. In general, any $\beta_n < 1$ indicates that the population of subjects in turn n underweights the information revealed by the history of others' actions relative to their private information. This is a plausible response to the belief that others can make errors in their decisions. To illustrate, Figure 7 shows sample plots of $\bar{\mathbf{z}}_\mathbf{n}$ with $\alpha_n = 0$ and $\beta_n = \beta$ for all n and differing values of $\beta \in [0,1]$.

[Figure 7 here]

Table 5 summarizes the econometric results[10] and compares them with the results of Çelen and Kariv (2004a). Note that under imperfect information both the $\hat{\alpha}_n$ and $\hat{\beta}_n$ coefficients are not significantly different from zero in all turns. Thus, we conclude that under imperfect information overall follow-own-signal heuristic outperforms Bayes' rule as a predictor. Under perfect information, in contrast, although in Bayesian terms, subjects assign too much weight to their own information and too little weight to the public information, they gradually increase their confidence in the information revealed by the history of actions taken before them, as $\hat{\beta}_n$ exhibits an upward trend showing that over time subjects tend to adhere more closely to Bayesian updating.

[10] GLS random-effects (mixed) estimators and robust variance estimators for independent data and clustered data yield similar results.

[Table 5 here]

In sum, over time, while under perfect information, the information revealed by the history of actions is relied upon more and subjects become increasingly likely to imitate their predecessors, under imperfect information subjects do not tend to rely more on the information revealed by the predecessor's action.

5 Discussion

The decision problems under perfect and imperfect information differ radically. The dissimilarities have two related sources. First, under perfect information any history of actions is shared as public information by all successors and, thus, everyone can infer perfectly what each of her predecessors has observed. Under imperfect information, in contrast, all learn only from their immediate predecessor's action. As a result, no subset of the history of actions is shared as public information, and thus everyone draws different inferences about what predecessors have observed.

Second, while under perfect information the valuable information revealed by the frequency of past actions is available, under imperfect information no one can tell if her predecessor is a deviator or an imitator. Thus, Bayesian inference induces a probability measure over all possible histories conditional on the immediate predecessor's action, such that the information embedded in the history is suppressed in a way that gives a significant weight to the event in which all predecessors acted as the immediate predecessor did. Put differently, because Bayesian individuals attempt to capture the content of all predecessors' signals by using their immediate predecessor's action, they become increasingly likely to imitate.

The pattern of our experimental results suggests two important conclusions. The first deals with the group behavior. Under imperfect information, herd behavior develops much less frequently than under perfect information, and even less frequently than the theory predicts. The second inference, which narrows the possible explanations for the first observation, is related to the individual behavior. The difference in group behavior is in fact a compositional difference in the individual behavior, representing the distribution of decisions over our concurring and contrary categories and is not attributable to differences in the persuasiveness of predecessors' actions once there is the desire to confirm.

Our results under imperfect information suggest that the individual behavior is less consistent even with the generalized Bayesian behavior. In view of these findings, one may ask how can we reconcile this with the conclusions reached in Çelen and Kariv (2004a) under perfect information. Obviously, in our informationally constrained environment, it is understandable that subjects are less likely to be able to act rationally. To organize our experimental data theoretically and to put the observed behavior into perspective, we use a modification of the Bayesian model, which provides a framework that enables us to understand the differences in individual behavior under perfect and imperfect information.

We proceed as Smith and Sørensen (2000) did (for detailed analysis, see their MIT working paper with the same title) and pursue a modification of the original model that abandons the assumption of common knowledge of rationality. To be precise, assume that a fraction p of individuals are noisy, and that whether an individual's behavior is noisy is unobservable by others and that the noise is distributed independently across individuals. We assume two forms of noise, which are at the opposite extreme, either noisy individuals take actions randomly by setting their cutoffs at either -1 or 1 with equal chance (Noise I), or noisy individuals ignore history and make decisions solely on the basis of private information, by simply setting cutoffs at zero (Noise II). As such, the actions of noisy individuals of the first type do not reveal any information to successors, whereas, put side by side with a rational individual, a noisy individual of the second type reveals more of her private information.

After adding noisy individuals of the first type to the imperfect-information model, simple calculations show that the adjusted cutoff dynamics of rational individuals follow the process[11]

$$\hat{\theta}_n = \begin{cases} -\frac{1+\hat{\theta}^2_{n-1}}{2} + p\frac{1+\hat{\theta}^2_{n-1}}{2} & \text{if} \quad x_{n-1} = A, \\ \frac{1+\hat{\theta}^2_{n-1}}{2} - p\frac{1+\hat{\theta}^2_{n-1}}{2} & \text{if} \quad x_{n-1} = B \end{cases} \tag{13}$$

and after adding noise of the second type, the cutoff process is given by

$$\hat{\theta}_n = \begin{cases} -\frac{1+\hat{\theta}^2_{n-1}}{2} + p\frac{\hat{\theta}^2_{n-1}}{2} & \text{if} \quad x_{n-1} = A, \\ \frac{1+\hat{\theta}^2_{n-1}}{2} - p\frac{\hat{\theta}^2_{n-1}}{2} & \text{if} \quad x_{n-1} = B \end{cases} \tag{14}$$

where $\hat{\theta}_1 = 0$.

In the language of Section 3, it is straightforward to see that with both forms of noise the imitation set $[\overline{\theta}_n, \underline{\theta}_n)$ is still increasing in n, i.e., $[\overline{\theta}_n, \underline{\theta}_n) \supset$

[11]Since both forms of noise are symmetric around zero, after adding noise the Bayesian inference of rational individuals is still symmetric, i.e., $\overline{\theta}_n = -\underline{\theta}_n$.

$[\overline{\theta}_{n-1}, \underline{\theta}_{n-1})$, and thus, over time, it is more likely that imitation will arise. However, in contrast to the result of the noise-free model, the imitation set $[\overline{\theta}_n, \underline{\theta}_n)$ never reaches $[-1, 1]$ for any n. Put differently, $\{\overline{\theta}_n\}$ and $\{\underline{\theta}_n\}$ are bounded away from -1 and 1 respectively.

Figure 8 illustrates the law of motion for $\overline{\theta}_n$ (the cutoff of a rational individual when the predecessor took action A) with both forms of noise. Note that the adjusted cutoffs, (13) and (14), are far above their noise-free counterparts (6), meaning a relative predisposition of subjects to follow their private information. Furthermore, over time, the gap between the noise-free and adjusted cutoffs increases, which suggests that the Bayesian solution, as given by cutoff strategy (1) and (2), does not predict behavior in the laboratory. Hence, noise has a dramatic impact as compared to the noise-free model there is much less chance that a rational individual will imitate her predecessor. This noise possibly explains why we find behavior in the laboratory under imperfect information inconsistent with the predictions of noise-free theory.

[Figure 8 here]

A characteristic of the imperfect information model with these two extreme forms of noise is instability that is more episodic, because a single action is necessarily less informative. Put differently, much less information is accumulated, rational individuals are not as likely to imitate their predecessors as in the noise-free model. Consequently, we observe fewer periods of uniform behavior and switches that are more frequent than the theory predicts.

In contrast, with both forms of noise, we next show that under perfect information individuals gradually increase their confidence in the information revealed by the actions of others. After adding noisy individuals of the first type to the perfect-information model, the rational-cutoff dynamics follow the stochastic process

$$\hat{\theta}_n = \begin{cases} \frac{-1+\hat{\theta}_{n-1}}{2} + p\frac{1+\hat{\theta}_{n-1}}{2} & \text{if} \quad x_{n-1} = A \\ \frac{1+\hat{\theta}_{n-1}}{2} + p\frac{-1+\hat{\theta}_{n-1}}{2} & \text{if} \quad x_{n-1} = B \end{cases} \tag{15}$$

where $\hat{\theta}_1 = 0$.

One can easily verify that the contributions of rational individuals to the learning process are not lost in the background noise. As a result, over time, the public information is given more weight, and rational individuals are more and more likely to join a herd. Moreover, a deviation from herd behavior has less impact, as it is likely to be noisy. Hence, this simple modification yields a noise-adjusted cutoff process that has the same trend as the noise-free cutoff process (8).

Alternatively, one can assume the second form of noise. Put side by side with a rational individual, under perfect information such a noisy individual reveals more of her private information by joining a herd but less by avoiding it. After adding such noise to the model, the cutoff process is given by

$$\hat{\theta}_n = \begin{cases} \frac{-1+\hat{\theta}_{n-1}}{2} + p\frac{\hat{\theta}_{n-1}}{2} & \text{if} \quad x_{n-1} = A \\ \frac{1+\hat{\theta}_{n-1}}{2} + p\frac{\hat{\theta}_{n-1}}{2} & \text{if} \quad x_{n-1} = B \end{cases} \tag{16}$$

where $\hat{\theta}_1 = 0$.

It is straightforward to verify that now the noise-adjusted cutoff process easily escapes the interval $(-1, 1)$, meaning that an informational cascade occurs. By making more of their private information public, the noisy individuals trigger a cascade stage, that otherwise never would start. Note, however, in this modified state of affairs the actions of those who ignore the history remain informative even during a cascade, and thus it is not an absorbing state. Figure 9 compares the noise-free cutoff process (8) and the adjusted processes for both forms of noise, (15) and (16), when all take action A.

[Figure 9 here]

To conclude, clearly, some complex multilateral mixture of bounded rationality and limits to the rationality of others can best characterize the nature of behavior. However, taken as a whole, generalized Bayesian behavior that is properly modified to take these traits into account permits successful prediction of the subjects' behavior under perfect information. Under imperfect information, in contrast, behavior is not consistent even with this generalization of Bayesian behavior.

6 Concluding Remarks

This paper tests an imperfect-information observational learning model that theoretically yields behavior quite distinct from and in some ways more extreme than that in the perfect-information model. Furthermore, using a setup with continuous signal and discrete action, along with a cutoff elicitation technique, enables us to examine how well Bayes rationality approximates the actual behavior observed in the laboratory.

Our results can be summarized briefly as follows. First, herd behavior is much less frequent under imperfect information than under perfect information, and even less frequent than the theory predicts. Second, the difference from the prediction of the theory is in fact a compositional difference representing the distribution of decisions over concurring and contrary categories and is not attributable to differences in how persuasive predecessors' actions are, once they are followed. In fact, in the subset of concurring decisions, there is a substantial degree of conformity with the theory.

The experiment tests the robustness of results derived in the perfect-information version of the observational learning experiments, and generate sharp and suggestive predictions. It is natural to ask about the robustness of the results when the number of most recent actions that a subject observes exceeds one. Our analysis does not properly address this issue since for any observation of histories larger than one the structure of the decision rule is extremely involved. Clearly, further inferences based on the frequency of past actions can be obtained and individuals are then able to identify deviators and imitators. Whether an increase in the number of predecessors observed would lead to sharply different results is not clear as different information structures may lead to different outcomes. This remains a subject for further research.

References

[1] Allsopp, L. and J. Hey (2000) "Two Experiments to Test a Model of Herd Behavior." *Experimental Economics*, *3*(2), pp. 121-36.

[2] Anderson, L. (2001) "Payoff Effects in Information Cascade Experiments." *Economic Inquiry*, *39*(4), pp. 609-15.

[3] Anderson, L. and C. Holt (1997) "Information Cascades in the Laboratory." *American Economic Review, 87*(5), pp. 847-62.

[4] Banerjee, A. (1992) "A Simple Model of Herd Behavior." *Quarterly Journal of Economics, 107*(3), pp. 797-817.

[5] Bikhchandani, S., D. Hirshleifer and I. Welch (1992) "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascade." *Journal of Political Economy, 100*(5), pp. 992-1026.

[6] Bikhchandani, S., D. Hirshleifer and I. Welch (1998) "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades." *Journal of Economic Perspective, 12*(3), pp. 151-70.

[7] Chamley, C. (2003) "Rational Herds Economic Models of Social learning." Cambridge University Press.

[8] Chamley, C. and D. Gale (1994) "Information Revelation and Strategic Delay in a Model of Investment." *Econometrica*, *62*(5), pp. 1065-85.

[9] Çelen, B. and S. Kariv (2004a) "Distinguishing Informational Cascades from Herd Behavior in the Laboratory." *American Economic Review,* Forthcoming.

[10] Çelen, B. and S. Kariv (2004b) "Observational Learning Under Imperfect Information." *Games and Economic Behavior, 47(1),* pp. 72-86.

[11] Gale, D. (1996) "What Have We Learned from Social Learning?" *European Economic Review, 40*(3-5), pp. 617-28.

[12] Gul, F. and R. Lundholm (1995) "Endogenous Timing and the Clustering of Agents' Decisions." *Journal of Political Economy*, *103*(5), pp. 1039-66.

[13] Hung, A. and C. Plott (2001) "Information Cascades: Replication and an Extension to Majority Rule and Conformity-Rewarding Institutions." *American Economic Review, 91*(5), pp. 1508-20.

[14] Kübler, D. and G. Weizsäcker (2003) "Limited depth of reasoning and failure of cascade formation in the laboratory." *Review of Economic Studies*, *71(2)*, pp. 425-41.

[15] Lee, I. H. (1993) "On the Convergence of Informational Cascades." *Journal of Economic Theory*, *61*(2), pp. 396-411.

[16] Smith, L. and P. Sørensen (2000) "Pathological Outcomes of Observational Learning." *Econometrica, 68*(2), pp. 371-398.

Table 1: Data for rounds in which herd behavior arises

Session. round*	Action Length	Action and cutoff by turn								Sum of signals
		1	2	3	4	5	6	7	8	
1.7	*B*	*B*	*B*	*B*	*B*	*B*	*B*	*B*	*B*	-46.6
	8	5	-4	10	0	0	2	0	-2	
2.7	*A*	*A*	*B*	*A*	*A*	*A*	*A*	*A*	*A*	23.2
	6	-10	2	-5	2.4	-10	-10	0	-4	
3.10	*A*	*B*	*B*	*B*	*A*	*A*	*A*	*A*	*A*	12.6
	5	0	4	8	-10	-10	0	-8	0	
3.12	*A*	*B*	*A*	*B*	*A*	*A*	*A*	*A*	*A*	39.5
	5	10	-10	6.9	-8	0	0	-4	0	
4.5	*B*	*A*	*B*	*A*	*B*	*B*	*B*	*B*	*B*	-5.5
	5	0	2.5	5.6	7	-1	10	10	9	
4.10	*B*	*A*	*B*	*A*	*B*	*B*	*B*	*B*	*B*	-16.7
	5	-10	0	-8	1.7	0	0	10	10	
4.11	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	*A*	28.3
	8	-7.5	1	3	-10	-3	0	3.3	-5.2	
5.8	*A*	*A*	*B*	*B*	*A*	*A*	*A*	*A*	*A*	35.3
	5	0	7	5	5	1.4	2	-6.7	-1.2	

* - (Session.Round). For example, 1.11 is the eleventh round in the first session.

Table 2: Data for selected rounds from session 2

Action
Cutoff
Private Signal

Session/ round	1	2	3	4	5	6	7	8	Sum of signals
2.3	*A* 0 1.19	*A* -6 4.88	*A* -2 9.16	*A* -4 7.9	*B* -1 -9.83	*A* 6 7.97	*B* 10 7.46	*A* -5.5 6.34	35.1
2.4	*B* 0 -0.13	*B* -1 -7.21	*B* 10 5.12	*B* 4 1.33	*B* 0 -4.45	*A* -5 -4.25	*A* -8 -0.42	*A* -10 -9.66	-19.7
2.10	*A* 1 7.09	*A* -5 -2.41	*A* -5 -4.58	*B* 0 -3.14	*B* 10 4.68	*B* 4 1.74	*A* -3 2.56	*A* -10 7.77	13.7
2.11	*B* 0 -8.03	*B* -6 -8.49	*B* 1 -2.02	*B* -5 -6.23	*B* 10 4.84	*B* 10 8.78	*A* -5 3.77	*B* 0.6 0.45	-6.9
2.12	*A* 0 9.42	*A* -6 4.63	*A* -5 -3.43	*A* -10 6.06	*A* 0 8.15	*A* -4 0.58	*B* 5 3.72	*B* 8 2.1	31.2
2.13	*B* -1 -3.06	*B* 3 -4.18	*A* -5 7.32	*A* -4 8.38	*A* 5 8.3	*A* -10 0.14	*B* 2 -0.72	*B* 10 9.51	25.7
2.14	*B* 0 -3.4	*B* -5 -9.13	*A* -5 -1.9	*A* -6 -2.05	*A* -4 2.44	*B* 6 -5.55	*B* 10 5.53	*A* 4 7.06	-7.0

Key: - Cascade behavior.

Table 3: Summary of experimental results

	Imperfect Information	Perfect Information
Earnings	$18.8	$22.0
Herds*	8	28
% of Herds**	10.7	37.3
Incorrect Herds	0	1
Cascades	18	26
% of Cascades**	24.0	34.7
Overturns	234	173
% of Overturns***	44.6	32.9

* Herds of at least five subjects.
** Out of all 75 rounds.
*** Out of all 525 decision points excluding the first decision turn.

Table 4: Regression results

	Coef.	Std. Err.	t
Turn	0.19	0.126	1.522
FR	-0.48	0.617	-0.775
LR	-0.19	0.617	-0.312
Cons.	0.18	0.765	0.239

1. A regression of the transformed cutoffs on the decision turn at which this cutoff was set as well as dummies which take a value of one in the first (FR) and last (LR) five rounds in a session (# of obs.=525).
2. GLS random-effects (mixed) estimators and robust variance estimators for independent data and clustered data (data not independent within subjects but independent across subjects) yield similar results.

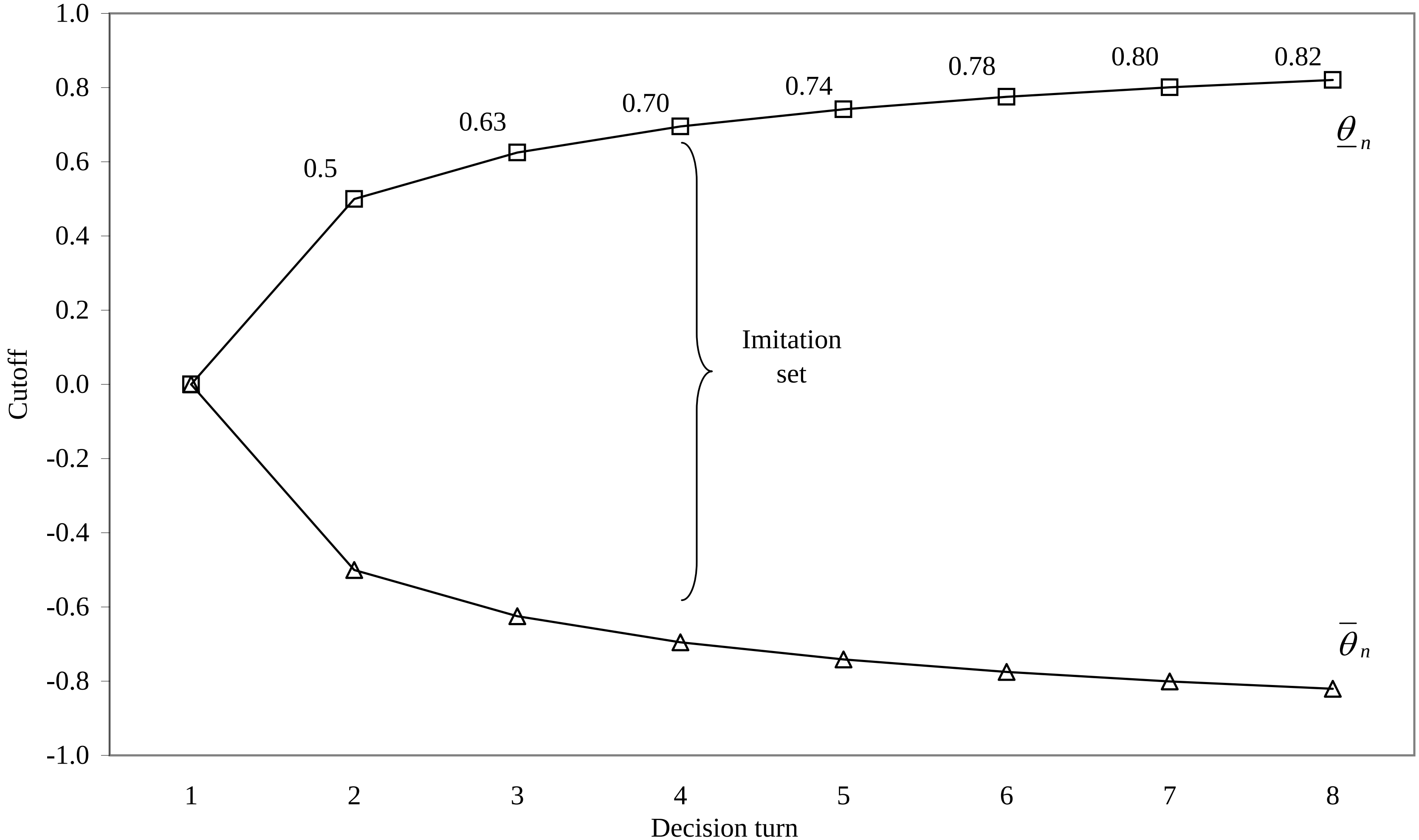
Table 5: The econometric results by turn

		Turn	2	3	4	5	6	7	8
		# of obs.	75	75	75	75	75	75	75
Imperfect information	$\hat{\alpha}$		-0.09 (0.06)	-0.10 (0.66)	-0.12 (0.72)	-0.57 (0.65)	-0.42 (0.67)	0.36 (0.67)	-0.56 (0.73)
	$\hat{\beta}$		-0.06 (0.12)	0.21 (0.13)	0.22 (0.14)	0.15 (0.13)	0.21 (0.13)	0.25 (0.13)	0.29 (0.15)
Perfect information	$\hat{\alpha}$		0.96 (0.46)	0.02 (0.56)	0.16 (0.56)	-0.02 (0.48)	0.39 (0.59)	-0.05 (0.63)	0.27 (0.67)
	$\hat{\beta}$		0.22 (0.09)	0.48 (0.07)	0.49 (0.07)	0.59 (0.06)	0.60 (0.07)	0.59 (0.08)	0.62 (0.08)

(Std. Err)

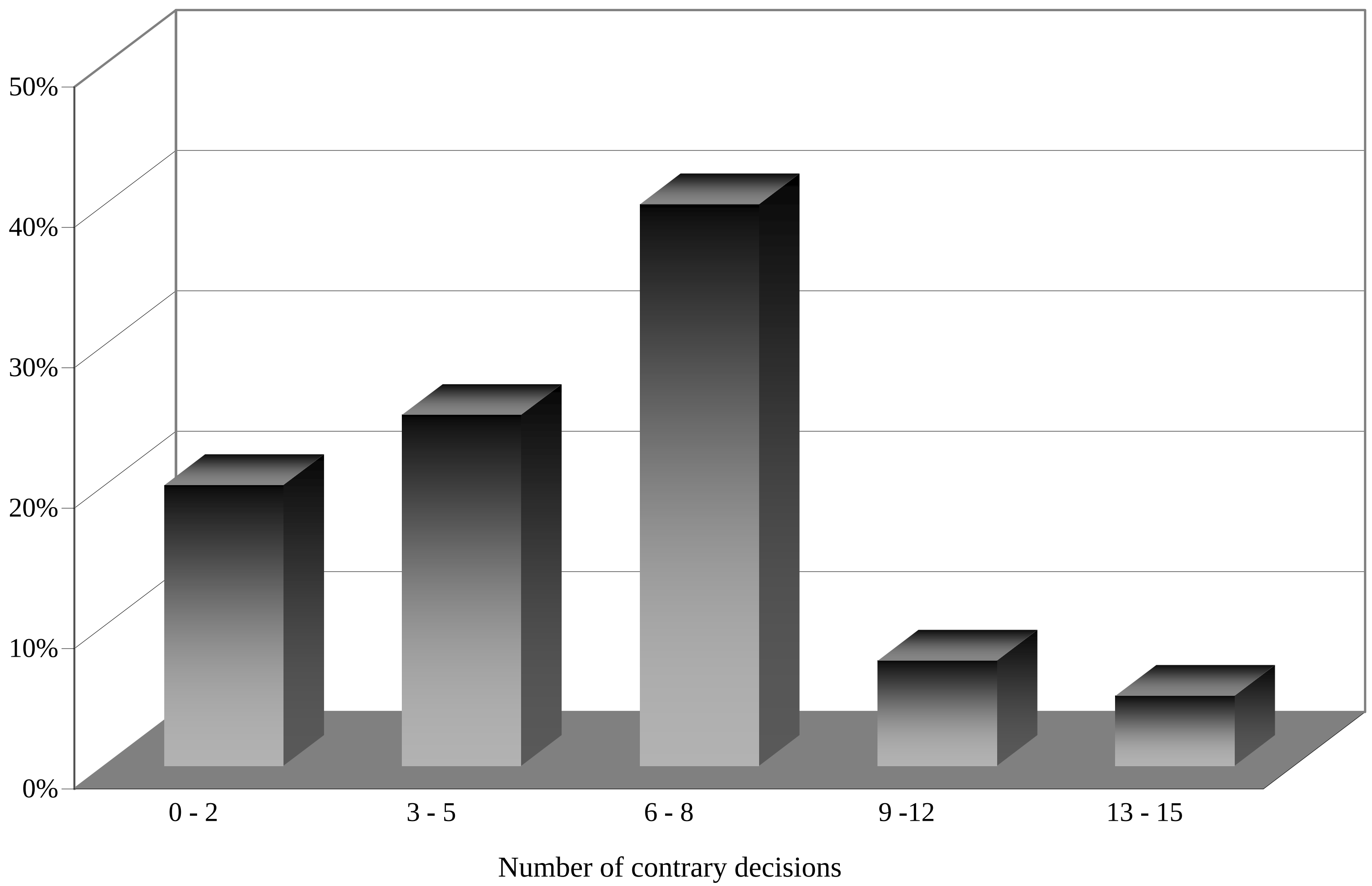
1). The econometric results under imperfect and perfect (Çelen and Kariv (2001b)) information.
2). Under imperfect information, both coefficients are not significantly different from zero in all decision turns, where under perfect information Betas exhibit an upward trend indicating, that over time subjects tend to adhere more closely to Bayesian updating.
3). GLS random-effects (mixed) estimators and robust variance estimators for independent data and clustered data yield similar results.

Figure 1: The process of cutoffs and imitation sets



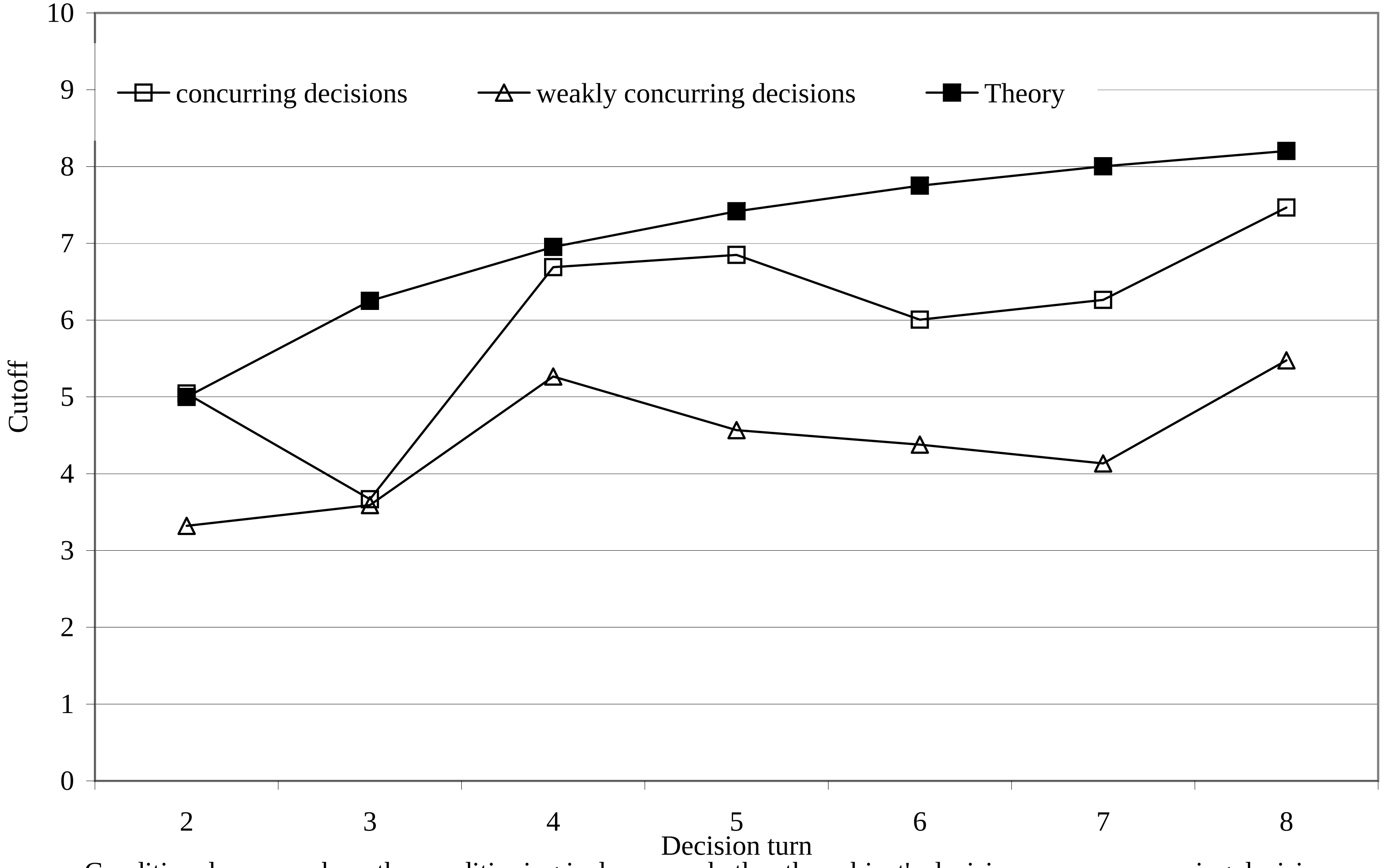
Imitation sets monotonically increase in n regardless of the actual history of actions, and thus, over time, it is more likely that imitation will arise.

Figure 2: The distribution of concurring subjects



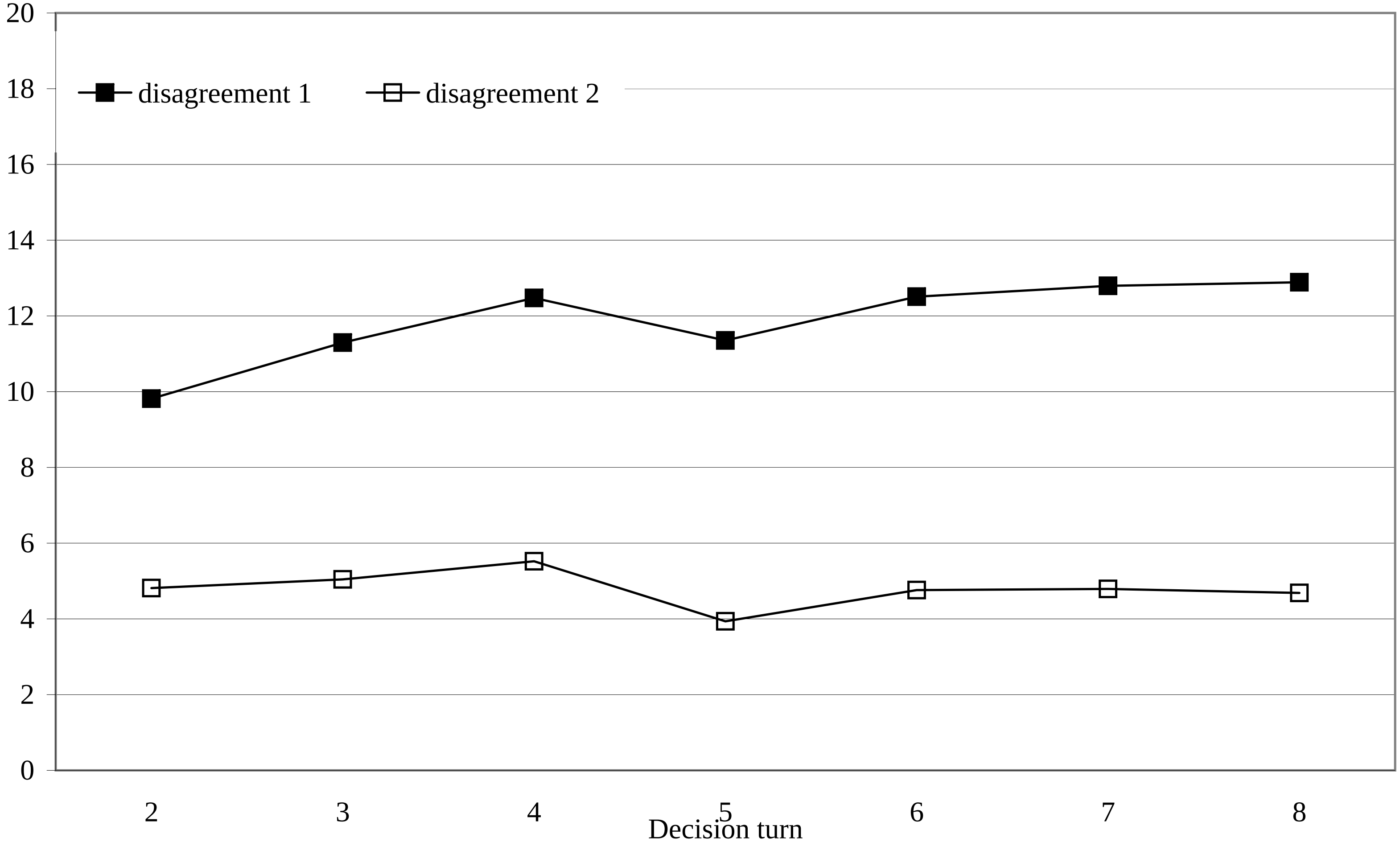
The percent of subjects who disagreed with the observed action in less than two rounds, three to five rounds and so on.

Figure 3: Mean cutoffs by decision turn in concurring and weakly concurring decisions



Conditional means where the conditioning is done on whether the subject's decision was a cooccurring decision.
In weakly concurring decisions, we include the neutral decisions in our means.

Figure 4: Strength of disagreement



The absolute value of the distance between the cutoff chosen and that which would be set if the subject acted according to the theoretical cutoff rule (disagreement 1), and between the cutoff chosen and zero (disagreement 2).

Figure 5: Unconditional mean cutoffs by decision turn

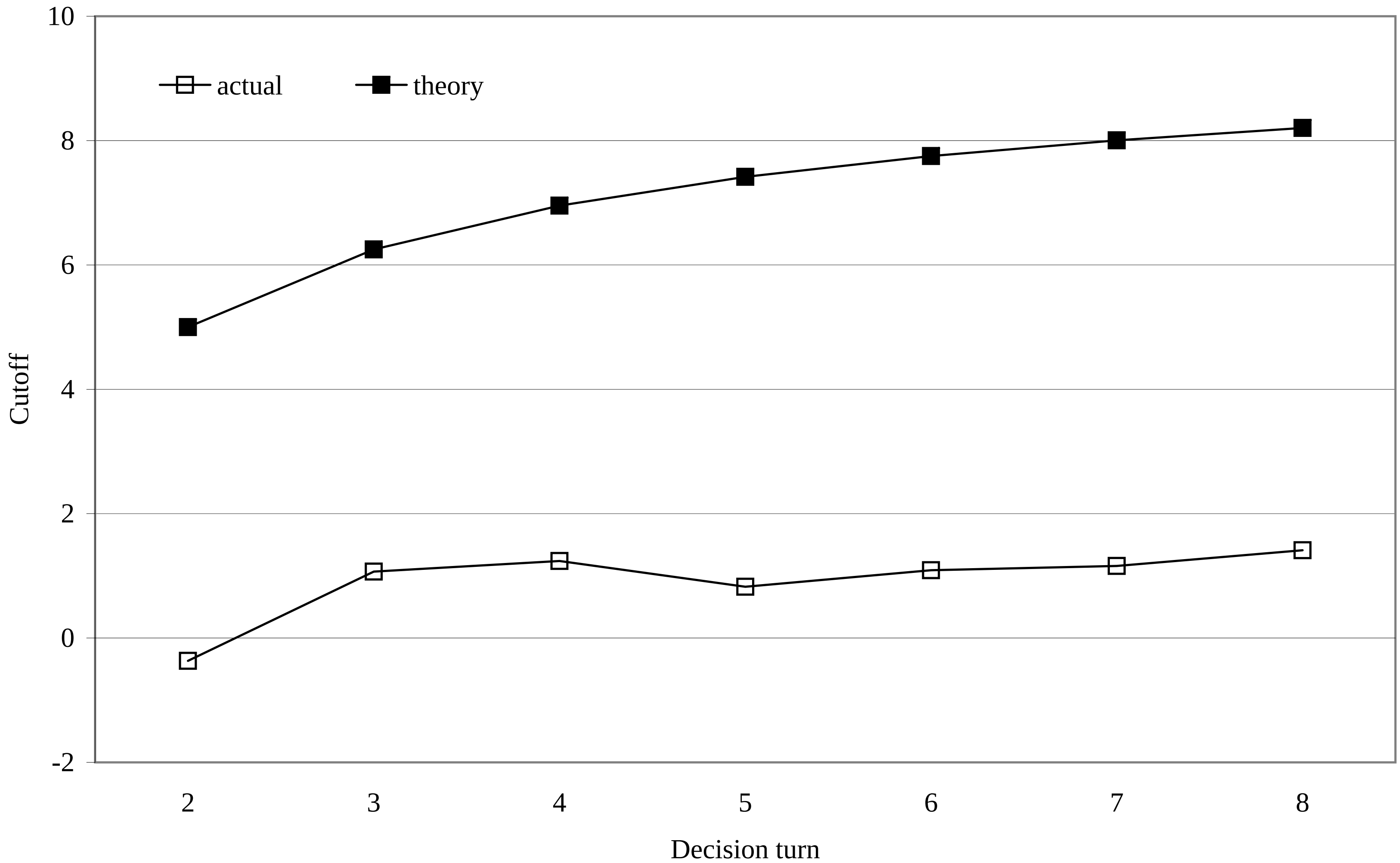
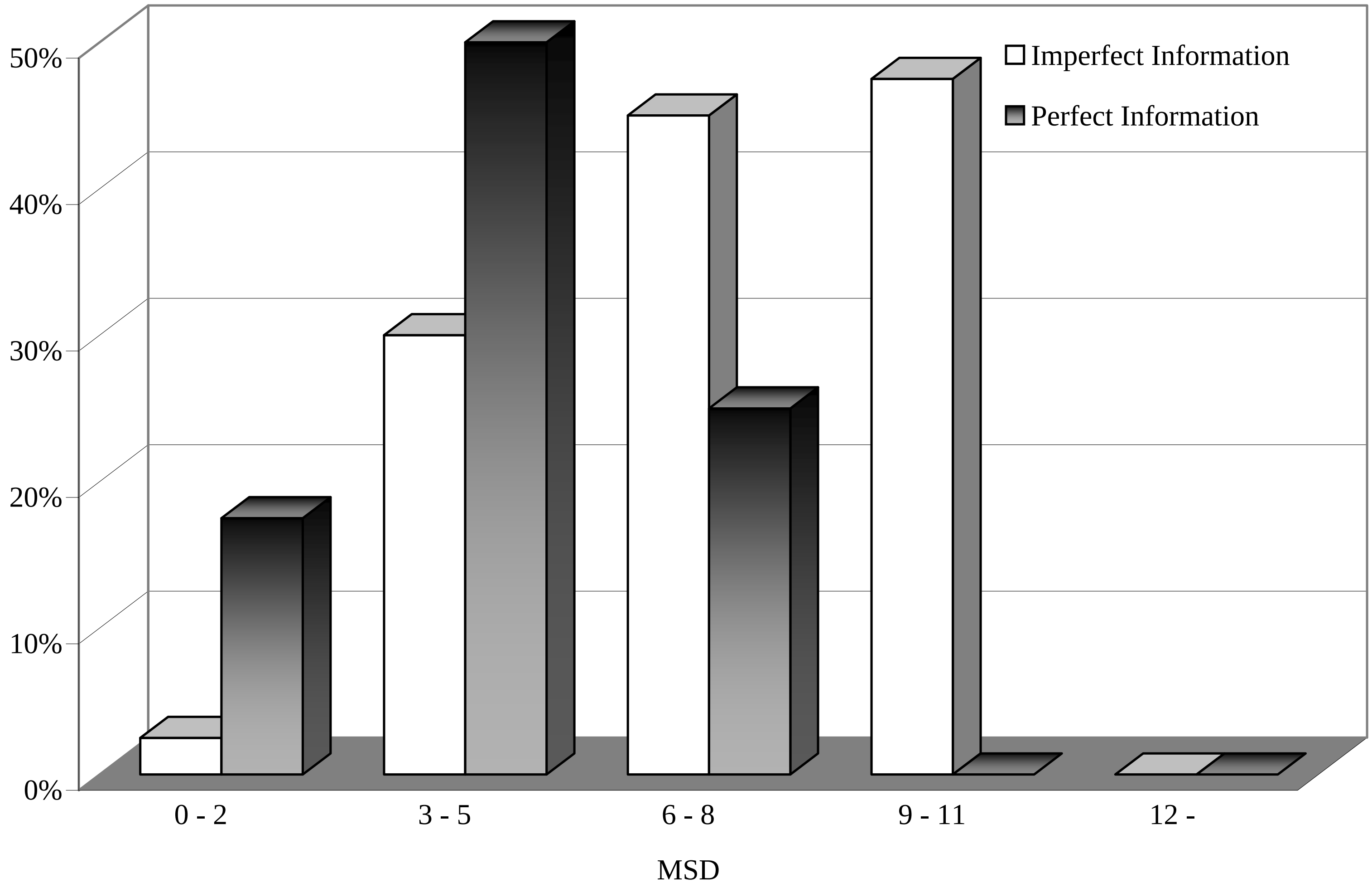
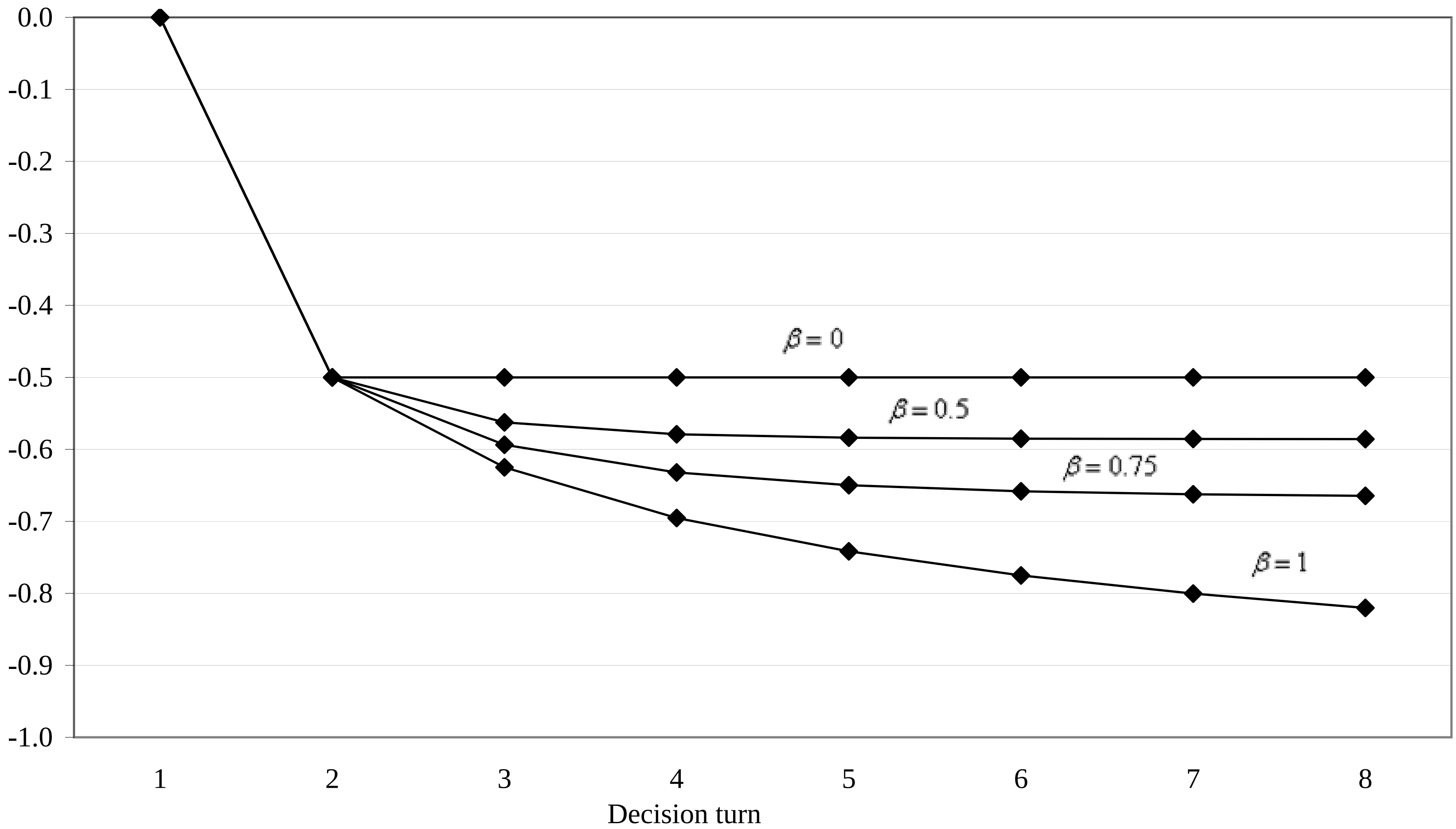


Figure 6: The distribution of subjects' MSD scores



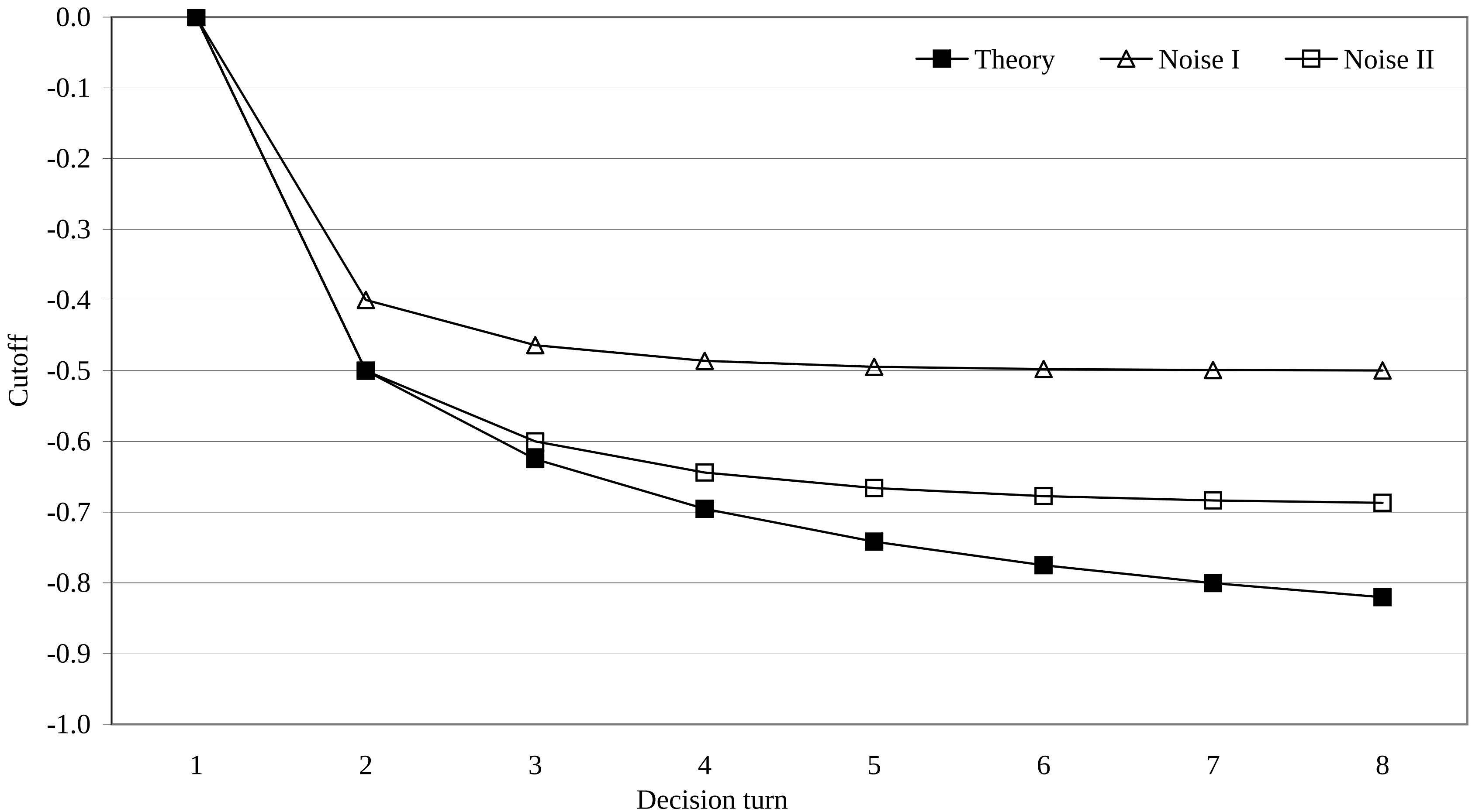
The histograms show that subject behavior is more consistent with the theory under perfect information as the distribution of MSD scores shifts to the left when calculated using the perfect information data.

Figure 7: Sample plots of the error-adjustment updating rule



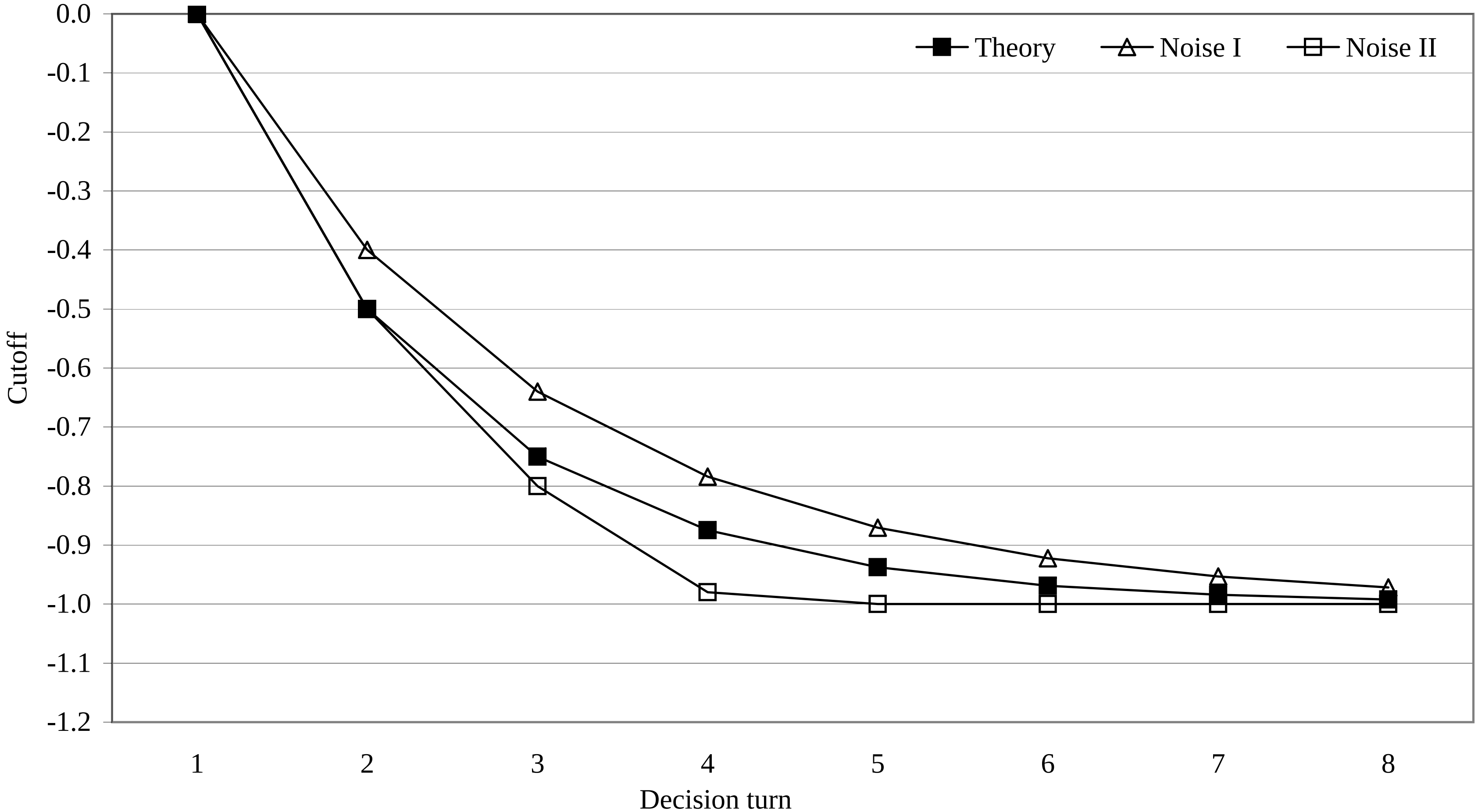
Sample plots of the error-adjustment updating rule with $\alpha_n = 0$ for all n and different levels of β. Note that when $\alpha_n = \beta_n = 0$ for all n, $\hat{\theta}_n = -0.5$ for any individual $n \geq 2$.

Figure 8: The sequences of cutoffs when all individuals choose action *A* under imperfect information
Two extreme forms of noise



With these two extreme forms of noise (p =0.2), individuals' cutoffs are far above their theoretical counterparts, meaning a relative predisposition of subjects to follow their private information.

Figure 9: The sequences of cutoffs when all individuals choose action A under perfect information
Two extreme forms of noise



With these two extreme forms of noise (p =0.2), individuals gradually increase their confidence in the information revealed by the actions of others and either case yields an adjusted cutoff process whose trajectory is similar to that of estimated cutoff process.