
Inexperienced and Experienced Players in an Oligopolistic Market Game with Minimal Information

ROSEMARIE NAGEL^a and NICOLAAS J. VRIEND^b
(^aUniversitat Pompeu Fabra, Barcelona, Spain and ^bQueen Mary and
Westfield College, University of London, UK. Email: n.vriend@qmw.ac.uk)

We consider an oligopolistic market game, in which the players are competing firms in the same market of a homogeneous consumption good. The consumer side is represented by a fixed demand function. The firms decide how much to produce of a perishable consumption good, and they decide upon a number of information signals to be sent into the population in order to attract customers. Due to the minimal information provided, the players do not have a well-specified model of their environment. Given a simple model of adaptive behavior, which we showed in a previous paper explained the data relatively well, our main objective here is to analyze whether experienced players behave differently to inexperienced players.

1. *Introduction*

Learning and adaptive behavior have recently become fashionable in economics. Conceptually, these two concepts are closely related, and they are frequently used as mere synonyms. In many models of *adaptive behavior* the agents adapt to their environment. If the players interact with each other, part of their environment will be adapting to their adaptation as well, and we get a coevolutionary process. Although such dynamics can be quite interesting, the bearing of many of these models of adaptive behavior is limited by the feature that putting the same players again in a similar starting situation would lead to qualitatively the same (expected) dynamics. What would be real *learning* is if the players adjust *the way* in which they adapt to their environment.¹ In this paper we will distinguish the two concepts carefully,

¹ Notice that these labels are somewhat arbitrary. In some sense, what we call 'learning' is just 'adaptive behavior' at a one step higher level. Different examples in the literature might use different names for these levels. But what seems to be useful is to distinguish these levels as such.

and we will measure the relative importance of each in an experimental oligopoly game.

In a previous paper (Nagel and Vriend, 1998) we considered the following oligopolistic market game. The players are identical competitors in the same market. In each period, they decide how much to produce of a perishable and homogeneous consumption good, and they decide how many advertising signals to send into the population in order to attract customers. The firms know the parameters of the production and signaling technologies, as well as the fixed price of the good. They also know that the consumers in this economy are simulated by a computer program. After every period, each firm observes only its *own market outcomes*. No further information about the environment is given. Given the minimal information, even a rational player is not in a position to maximize his profits using standard optimization techniques. Hence, instead of *deducing* optimal actions from universal truths, he will need *inductive* reasoning; proceeding from the actual situation he faces. That is, he must behave adaptively.

We used a simple two-step model to put the experimental data into perspective. The two-step model is based on the minimal information that the players actually got, while making only minimal assumptions about the agents' reasoning processes. The model consists of two simple processes: learning direction theory, which has been successfully applied in various experiments (e.g. Selten and Stoecker, 1986; Nagel, 1995); and the well-known method of hill climbing, also known as gradient method. As we will make clear below, these two steps share the following underlying principle. The players' own actions and outcomes in the most recent (two) period(s) give the player information about the direction in which he may find better actions. We will also use this two-step model to analyze the differences in actions and outcomes between the players. We found that the behavior of the players *on average* largely conforms to the predictions of the simple two-step model. Also the time pattern of the average behavior showed a good fit with the two-step model.

This two-step model describes how the players adapt to their circumstances. But our main objective in this paper is to know whether the players also learn to *adjust* the way in which they adapt to their circumstances. Therefore, we asked all players to return for a very similar oligopoly game with experienced players, and consider the following kinds of questions. Are there qualitative or quantitative differences in the adaptation process? Do they play more or less aggressively? Is there a substantive difference in behavior in the first period? Surprisingly, we cannot find a significant difference between inexperienced and experienced players with respect to the

TABLE 1. Notation and Parameter Values

Symbol	Meaning	Inexperienced	Experienced	Known
c	'marginal' cost production	0.25	0.27	yes
f	patronage rate satisfied consumers	0.56	0.63	no
g	price minus 'marginal' cost production	0.75	0.73	yes
k	'marginal' cost signaling	0.08	0.21	yes
m	no.of firms	6	6	no
n	no. of consumers	712	1422	no
N	total no. of agents	718	1428	no
p	price of the commodity	1.00	1.00	yes
Π	profit	—	—	own
q	demand directed towards a firm	—	—	own
Q	aggregate demand	—	—	no
s	no. of signals sent by a firm	—	—	own
—	maximum value for s	4999	4999	yes
S	aggregate no. of signals all firms	—	—	no
S_{-}	aggregate no. of signals other firms	—	—	no
V	value	—	—	no
x	sales	—	—	own
z	production	—	—	own
—	maximum value for z	4999	4999	yes
—	no. of periods	± 150	± 300	no

two-step model of adaptive behavior. Experienced players seem to adapt in the same way to their opportunities as the inexperienced players.

The paper is organized as follows. In Section 2, we present the oligopoly game, and in Section 3 the experimental setup. Section 4 presents a simple model of adaptive behavior. Section 5 contains an analysis of the behavior of the experienced players, and Section 6 concludes.

2. The Oligopoly Model

A fixed number of firms repeatedly interacts in an oligopolistic market. All firms are identical in the sense that they produce the same homogeneous consumption good, using the same technology (see below). Time is divided into discrete periods. At the beginning of each period, each firm has to decide how many units of the perishable consumption good to produce. The production costs per unit are constant, and identical for all firms. The production decided upon at the beginning of the period is available for sale in that period. The firms also decide upon a number of information or advertising signals to be sent into the population, communicating the fact that they are a firm offering the commodity for sale in that period. Imagine the sending of letters to addresses picked randomly from the telephone

directory. The costs per information signal sent to an individual agent are constant, and identical for all firms. The price of the commodity is fixed, invariant for all periods, and identical for all firms and consumers. The choice of the number of units to be produced, and the number of information signals to be sent is restricted to a given interval.

Table 1 gives the notation used throughout the paper. Superscripts will be used for the time index, and subscripts for the identity of a firm. In addition, Table 1 gives an overview of the parameter values used. The last column indicates whether the parameter value was known or not to the players. As we will explain below, in addition to these parameter values, the players did not know the functional form of the demand they faced.

Consumers in this economy are simulated by a computer program. In each period, when all firms have decided their production and signaling levels, consumers will be *'shopping'*, with each consumer wanting to buy exactly one unit of the commodity per period. In fact, the consumer side is represented by the following fixed, deterministic demand function:

$$q_i^t = \text{round} \left(\text{trunc} \left[f \cdot x_i^{t-1} \right] + \frac{S_i^t}{S^t} \cdot \left[1 - \exp \left(-\frac{S^t}{N} \right) \right] \cdot \left[n - \sum_{j=1}^m \text{trunc} \left(f \cdot x_j^{t-1} \right) \right] \right) \quad (1)$$

(I) + (IIa) · (IIb) · (IIc)

where

- I = demand directed towards firm *i* by patronizing consumers
- IIa = proportion of signals from firm *i* in aggregate signaling activity
- IIb = proportion of individuals reached by one or more signals
- IIc = number of 'free', i.e. non-patronizing, consumers
- IIa · IIb · IIc = demand directed towards firm *i* as a result of current signaling

In each period, a fixed fraction of the number of customers that were satisfied by a given firm during the last period will *patronize* that firm (part I of equation 1).² The remaining consumers who received at least one signal (part IIc multiplied by IIb) are split between the firms, according to the firms' signaling activity relative to the aggregate signaling in the market (part IIa). Notice that when all firms signal very little, not all consumers will be reached

² See also Keser (1992) for duopoly experiments with demand inertia.

by an information signal, implying that not all consumers will actually be present in the market. Hence, although all signaling has the form of informative advertising, and there is no persuasive signaling, one can distinguish two different effects. First, a business stealing effect, and second, an effect on the total demand in the market. In Vriend (1996), in a closely related model, we consider explicitly a process of sending, receiving and choosing individual signals, and show that this leads to a demand function faced by the individual firms that may be approximated by a Poisson distribution. The deterministic function given above equals the expected value of such a Poisson distribution. At the end of the period, all unsold units of the commodity perish. Notice that a firm cannot sell more than it has produced at the beginning of the period. Hence, a firm's profit in period t is given by:

$$\prod_i^t = p \cdot x_i^t - c \cdot z_i^t - k \cdot s_i^t, \quad \text{where } x_i^t = \min[z_i^t, q_i^t] \quad (2)$$

In order to obtain a theoretical benchmark, in Nagel and Vriend (1998) we derive the symmetric stationary optimal policy for a given player for any given period, assuming complete information about the demand function, and show that in the symmetric stationary equilibrium, the signaling level for an individual firm i in a given period t is given by:

$$s_i^t = \frac{g}{k} \cdot \left(\frac{m-1}{m^2} \right) \cdot n \quad (3)$$

and the production is simply equal to the demand thus generated since the demand function is deterministic (see equation 1). The numerical values of this equilibrium implied by the parameters of the model are a production level of 18 and signaling level of 927 for the inexperienced, and production at 232 with signaling at 687 for the experienced players.

Before presenting the experimental setup, and our analysis, let us explain a little bit more in detail the background of the specific oligopoly model used. First of all, we have a model with a fixed price, while the consumer side is simulated with a fixed deterministic demand function. In other words, we abstract from the process by which the price was determined, and from the determination of the market demand at that price level. Notice that this is perfectly compatible with a standard downward-sloping market demand curve. These two abstractions are made in order to focus on the behavior of the firms in a relatively stable environment. As shown in Nagel and Vriend (1998), since the firms influence each other's environment, the task of the

firm is already complicated. Once we have understood and structured the behavior in such an environment, we can further relax these assumptions.

Note, however, that there *are* many markets in which goods are sold at fixed prices (whether as a result of legislation, of vertically imposed restrictive practices, or of optimizing behavior of the sellers). And over the period for which prices are fixed, trading opportunities are usually uncertain. Doctors fees are regulated in many countries, the prices of books are fixed by publishers' cartels in a lot of countries, bread prices in Italy are determined centrally, newspapers usually sell for the same constant price at all stands. Nearly any retail operation has a posted price, invariant through time, orders for stock are placed in advance of knowing what demand will be, and stock outs are commonly faced. Levis jeans are sold that way. So are Seiko watches. McDonald's rarely runs out of food, but it does happen. A fascinating market where this is true is the motion picture industry (see De Vany and Walls, 1996). Admission price is set and does not vary over the run, films are booked at theaters before demand is known, customers cannot always get a seat at their preferred showing and must queue up or wait for another day. Also custom items, e.g. in the finished steel or medical industry, are usually produced to order, at previously posted prices.³ Clearly, a *complete* economic analysis would explain such legislation, restrictive practices or strategies, by which the prices are fixed, as well. Instead of explaining the price, our analysis focuses on the learning and adaptive behavior of the firms; and thus applies equally to all the possible ways in which these prices may have been determined. Given the price level, competition can take place along many dimensions. Competing for customers through advertising seems a fundamental one.

Advertising as a form of communication to identify potential trading partners is important in real markets. Some readers may wonder why we bother to consider this model with its explicit signaling structure, since in the real world *a consumer simply knows where to buy something*. Well-considered, such an observation gives strong empirical support for our approach. For it asserts that transactions do not take place in Walrasian central markets, or through anonymous random matching devices, but that, instead, market interactions depend in a crucial way on local knowledge of the identity of potential trading partners. Such information has to be communicated in one way or another. Most advertising in reality seems indeed to draw the buyers' attention to the fact that someone is selling something somewhere sometime. Notice that

³ The list of examples seems to suggest that this applies to almost all markets that are not centralized like stock markets or auctions. But notice that even the Israeli stock market collects bids and offers to find a clearing price only in the morning, and then that price is posted and remains *fixed* over the entire trading day. Further documentation concerning the (in)flexibility of prices can be found in Carlton (1989)].

the advertising signals do contain no information on prices, and that this conforms to what we usually observe in reality. Although Stigler (1961, p. 223) argued: 'From the manufacturer's viewpoint, uncertainty concerning his price is clearly disadvantageous. The cost of search is a cost of purchase, and consumption will therefore be smaller', daily experience suggests that sellers must perceive some advantage of *not* communicating price information.⁴ Advertising is not always in the form of signals sent directly to individual potential buyers. Agents may also hear from radio and television, or from friends about the newest shops in town, buyers may use the Yellow Pages to find a seller, or they may visit shops (recognizable as such) randomly, etc. While these possibilities would require slight modifications of the signaling framework used, they would seem to fit rather well into it. The essential characteristic of all these forms of signaling is that agents make information about their own type known to some other agents. The resources spent in these forms of market-making are enormous in a modern market economy.

How reasonable is it to assume that firms do not observe the decisions and outcomes of the other firms, and in particular their signaling levels? As we will show below, the only strategic variable to compete directly with the other firms in this model is the signaling activity. In reality, competition takes place along many dimensions. Quantity and production capacity are obvious ones. Product differentiation is another one. The quality of a good depends upon many aspects, e.g. a warranty, add-ons like frequent flyer miles or after-sales service. Firms also compete using entry deterring and other restrictive practices, by their choice of technology, location or the timing of new product lines. Further competitive variables are the firms' R&D decisions (including marketing research), and their efforts to build up a reputation. The bottom line of this partial list is that it seems more than plausible that for some of these dimensions the information that an individual firm has about its competitors is far from complete. Assuming that firms do not observe the level of their competitors' signaling activity in our simple model is a first approximation of this fact.

⁴ Not only do sellers usually advertise without communicating prices, they often furnish shopwindows, and sometimes even have shelves, without prices. A buyer who incurs costs to get such information, e.g. by making a phone call, can often expect a rude answer or sometimes no answer at all, and a buyer who writes down prices in a supermarket takes some risk of being threatened. And even when advertising signals appear to contain information on prices, this is not necessarily meant to be as informative as it appears. In an interesting recent paper Lazear (1995) analyzed the practice of using price signals in advertising as a bait to attract customers, in order to switch to different products with different prices instead when customers visit their shop.

3. *The Experimental Setup*

We conducted two series of experiments in the computerized experimental laboratory at the University of Bonn—one with inexperienced and one with experienced players. In the first series, we organized 13 sessions with inexperienced players. In each session, six firms were competing in one market, for a total of 78 players. After finishing that first series, we asked all players to return for a very similar oligopoly game with experienced players in order to test whether the players learn over time to adjust the way in which they adapt to their circumstances. In this second series with experienced players we organized five sessions, with again six firms per session, for a total of 30 players. In this paper we will focus on the behavior of the experienced players; in particular comparing it with the behavior of the inexperienced players. A more extensive analysis of the behavior of the inexperienced players can be found in Nagel and Vriend (1998).

Most players came from various departments of the University of Bonn. Players sat in front of personal computers, and could not observe the screens of other players. Figure A1 in the appendix presents an example of a computer screen viewed by a given player. In the sessions with inexperienced players we had played ~ 150 periods, whereas the sessions with experienced players lasted ~ 300 periods. There was no time limit for the participants' decisions. Each player got a fixed 'show-up' fee. In addition, each player was paid according to the total profits realized by his firm. Losses realized were subtracted from the 'show-up' fee. The total payoff for an individual player was given by: $\alpha + (\alpha/2000) \times (\text{points realized})$.⁵ Observe that bankrupt players had lost their 'show-up' fee, and hence got nothing. Each session lasted about 2.5 hours, and the average payment over the 30 players was DM 53.89 ($\sim \$35.50$).

All sessions with experienced players were independent from each other with respect to the experience of the players in the first series. That is, whenever more players from a certain session with inexperienced players participated again, they would stay together and not be split over different sessions with experienced players.

In sessions with experienced players, there are two options for choosing the parameter setup: either using the same setup as in the preceding session, or a different parameter setup. The problem with using the same parameter setup for the experienced players is that a player might have found a good strategy

⁵ The value for α was DM 25 in session 1, DM 7 in session 2, and DM 13 in sessions 23–25, giving an average α of DM 14.2. The values for α were varied in advance in an effort to keep average payoffs at a level of \sim DM 15/h.

by chance during the first experiment, without having learned anything general as to how to play.⁶ As we are interested not in specific actions for the game considered but in adaptive behavior in this type of environment, we chose the second option.⁷ Thus, players in the second series were experienced in the sense that they might have learned something about the general *structure* of the game, and about their adaptive behavior in such an environment.

We now sketch the information that was available to the individual players, distinguishing technology, market and experience factors. The appendix presents the instructions given to the players, and Table 1 above summarizes which parameter values were known and which not.

3.1 The Technology

The players knew that they were identical firms, producing the same homogeneous consumption good, using the same technology. Both the production and signaling technology were common knowledge, and the same applies to the price of the commodity. About the fact that the choice interval for production and signaling was limited, the players was told that ‘this is due only to technological restrictions, and has no direct economic meaning’.

3.2 The Market

The players were told that consumers in this economy are simulated by a computer program. They did *not* receive the specification of the demand function (equation 1), and they did not know the number of competing firms,⁸ the number of consumers or the parameter value of the demand inertia. Instead they were given the following general picture of the consumption side. Each consumer wants one unit of the commodity in every period. Hence, in each period, a consumer has to find a firm offering the commodity for sale, and that firm should have at least one unit available at the moment he arrives. The participants were given two considerations concerning consumers’ actions. First, a consumer who has received an information signal from a firm *knows* that this firm is offering the commodity for sale in that period. Second, consumers who visited a certain firm, but found only empty shelves might find that firm’s service unreliable. On the other hand, a consumer who

⁶ This problem may be mitigated somewhat by giving each experienced player all the statistics of the first series before the second series starts.

⁷ See Table 1 above for the two parameter setups used.

⁸ In all sessions with experienced players there were six players at the same time in the lab, but players did not know how many parallel sessions were going on.

succeeded in buying one unit from a firm might remember the good service, and might be more likely to come back. Participants was also told ‘experience shows that, in general, the demand faced by an individual firm is below 1000’.

3.3 Experience

At the end of the period, each firm observes only its own market outcomes, and never the actions and outcomes of the other players. Each firm knows the demand that was directed to it during the period, how much it actually sold, and its profits for that period. Sometimes the market outcomes are such that a firm makes a loss. A firm making cumulative losses is informed about these. Each firm faces a known upper limit for the total losses it may realize. A firm exceeding this limit is declared bankrupt, with the participant removed from the session. This was stated before the experiments begun. The players did not know the number of periods to be played, but they knew that the playing time would be ~ 2.5 hours. Given this minimal information, a player is not in a position to maximize his profits on the basis of a well-specified demand function. In other words, he finds himself in a ‘large world’, and must behave adaptively. During the instructions before the games, some players felt uncomfortable with so much ‘mist’, and most questions attempted to get more knowledge about the environment. The usual answer to those questions was ‘you just don’t know’.

4. *A Simple Model of Adaptive Behavior*

In case of complete information, the only choice variable for a firm would be the number of signals to be sent, whereas production should be simply adjusted to the demand generated by these signals. This suggested a two-step decision problem for the players in our experiment. The *first* step concerning the *number of signals* to be sent, and in the *second* step adjusting the *production level* to the level of the demand generated. We will first consider this second step.

4.1 Production: Learning Direction Theory

Given the demand generated by a players’ signals sent in the current and previous periods, the production level that would yield the highest profit would be equal to this demand. We conjecture that the players use a simple algorithm to achieve this. This is sometimes known in the experimental literature as *learning direction theory* (e.g. Selten and Stoecker, 1986; Nagel,

TABLE 2. Predictions: Learning Direction Theory

If	Then
(1) $\text{production}_t < \text{demand}_t$	$\text{production}_{t+1} \geq \text{production}_t$
(2) $\text{production}_t > \text{demand}_t$	$\text{production}_{t+1} \leq \text{production}_t$
(3) $\text{production}_t = \text{demand}_t$	NA

1995). In our game, this direction learning mechanism can be applied as follows. If a firm faced more demand than it had produced, it *knows* that a higher production level would have led to higher profits. And if a firm faced less demand than it had produced, it *knows* that a lower production level would have led to higher profits. Therefore, in our model learning direction theory would lead to the *predictions* given in Table 2. Notice that if production and demand were equal, the theory does not predict the direction of the change in production. Remember that, given the two-step model (setting signals and adjusting production), these predictions are for a *given* demand level. Clearly, as we will analyze below, the demand depends upon the signaling level. Therefore, here we only consider those cases in which the players did not move into the opposite direction with their signaling level in order to induce a demand change.⁹

Learning direction theory makes the prediction on the basis of the last period. The reasoning of the players is supposed to be boundedly rational in that it only considers what *would have been* a better action, i.e. *ex post*. The demand was generated by a fixed deterministic demand function, but since this was not known to the players, there was subjective uncertainty. Hence the problem for the players is not so much to maximize what their profit could have been (*ex post*), but to maximize their expected profits (*ex ante*). If demand is uncertain, and rationing is not all-or-nothing, some overproduction may be profitable. That is, the production that maximizes expected profits may be higher than the expected demand. We predict the players to recognize this in our experiment, and hence expect a bias towards ‘overproduction’ relative to the predictions of learning direction theory.

Learning direction theory is based on a simple reasoning process, which is in essence an error-correction mechanism. A player needs enough knowledge and understanding about the structure of his environment to see immediately in which direction his own previous action was off-target. He corrects his action parameters on the basis of some error function representing a measure of the distance between the ‘target’ action and his actual action. Whereas with

⁹ That is, if an increase in production is predicted there should be no decrease in signaling, and vice versa. This condition was satisfied in 63% of the cases over all experiments.

learning direction theory a player is supposed to understand himself where better actions would have been found, the learning mechanism as such is similar to the one used in so-called supervised learning algorithms, where it is the supervisor who tells a player after each trial what the 'target' action was.

In many cases, however, it might be that a player does not know which is the 'correct' (i.e. the best possible) action, or how much it differed from such a target, not even afterwards. Often, there is only a notion of what a player should accomplish plus a success measure of his performance. For example, a player has to generate profits or utility, mapping an observed state (input) to actions (output), where the measure of success is simply the amount of profits or utility. In such cases the player could learn through 'reinforcement' (e.g. Bush and Mosteller, 1955; Cross, 1983; *Machine Learning*, 1992; Roth and Erev, 1995). Reinforcement learning is based on two principles. First, players try actions. Second, actions that led to better outcomes in the past are more likely to be repeated in the future. The players do not need any understanding about the structure of their environment. Perhaps they do not even know what payoffs are, but they recognize it when they receive it. Sometimes error-correction mechanisms as sketched above are also called 'reinforcement learning'. As Barto *et al.* (1983) point out, that would be misleading. Error-correction mechanisms are not based on a relative assessment of *consequences* of the player's actions, but *only* on knowledge (of the supervisor) of both the correct and actual output. This does *not* involve feedback that passes through the player's environment.¹⁰

Notwithstanding these fundamental differences, there have been some recent attempts (see Stahl, 1996; Camerer and Ho, 1997) to combine the two. Starting with reinforcement learning, one could assume that a player does not only update the probability of choosing a certain action on the basis of the payoff actually realized, but that he also has enough knowledge about the structure of the game (similar to the assumption made with learning direction theory) to reason what the payoffs would have been for actions not actually chosen. On the basis of such reasoning, a player could use these virtual payoffs for virtual reinforcement of those actions. As this will imply a gradual movement towards the optimal action, it might seem *as if* such a player applies learning direction theory. One of the problems with this virtual reinforcement approach is that sometimes a player cannot know what payoff he would have obtained by an action not actually chosen. Applying virtual reinforcement in such cases would lead to a biased learning process (see Vriend, 1997).

¹⁰ See Vriend (1994) for a more extensive discussion of these issues.

TABLE 3. Predictions: Hill Climbing

If	Then
(1) $\text{signaling}_t < \text{signaling}_{t-1}$ and $\text{payoff}_t < \text{payoff}_{t-1}$	$\text{signaling}_{t+1} > \text{signaling}_t$
(2) $\text{signaling}_t < \text{signaling}_{t-1}$ and $\text{payoff}_t > \text{payoff}_{t-1}$	$\text{signaling}_{t+1} < \text{signaling}_{t-1}$
(3) $\text{signaling}_t > \text{signaling}_{t-1}$ and $\text{payoff}_t < \text{payoff}_{t-1}$	$\text{signaling}_{t+1} < \text{signaling}_t$
(4) $\text{signaling}_t > \text{signaling}_{t-1}$ and $\text{payoff}_t > \text{payoff}_{t-1}$	$\text{signaling}_{t+1} > \text{signaling}_{t-1}$
(5) $\text{signaling}_t = \text{signaling}_{t-1}$ or $\text{payoff}_t = \text{payoff}_{t-1}$	NA

4.2 Signaling: Hill Climbing

As said above, the adaptation of the production level is assumed to take place for a given demand level. Since this demand is generated eventually by the signals sent, it is time to turn to an analysis of the number of signals sent. Learning direction theory cannot predict much with respect to signaling since in many cases a player cannot deduce the optimal signaling level from the demand observed. Depicting a player's opportunities as a hill, the objective of a player is to find the top of such a hill. A problem for a player is that he does not know what his hill looks like, and the hill may be changing all the time. A simple way to deal with this problem would be to start walking in some direction; if one gets a higher payoff one continues from there, otherwise one goes back to try another direction. Eventually one should reach a top. We conjecture that the players' adaptive behavior in signaling space can be described by such a *hill climbing*, or gradient, algorithm.¹¹ A deterministic variant of hill climbing would give the *predictions* presented in Table 3.

In our experiment there is one complication with this hill climbing. The hills may change over time, even when the actions of the other players are constant. Therefore, given the *dynamics* of the demand generated by the signals sent and the patronizing customers, a player should look further ahead than his immediate profits only. Players could boost their immediate profits by signaling very little, i.e. by eating into their pool of customers. But future profits are adversely affected by this. The cause is that of all the customers satisfied in a given period, some fraction will come back 'for free' in the next period, i.e. without the need to send them a signal. A firm can also forego some current profits by investing in the buildup of a pool of customers. The higher the current sales level, the better the firm's future profit opportunities. Hence, when considering the question how well a firm performed in a given

¹¹ See also Bloomfield (1994), Kirman (1993), Roberts (1995) and Merlo and Schotter (1994). The reason that we do not use reinforcement learning (as, e.g., in Roth and Erev, 1995) is that it is not clear how to reduce the choice set of the players (e.g. Holland *et al.*, 1986). Much more progress needs to be made here.

period, one should not only look at its immediate profits, but also at the change in its current sales level. The value of serving additional customers now (besides the immediate profits) is the profits that can be extracted from them in later periods.¹² Since patronizing customers come back ‘for free’ (without needing a signal), the profit margin for those customers will be the price minus the unit production costs of the commodity. Formally, the lookahead payoff in a given period is:

$$\Pi + \Delta x \cdot (p - c) \cdot \sum_{t=1}^{\infty} f^t$$

We will consider both the basic hill climbing variant, in which the players go myopically for their immediate profits only, and the variant in which the players climb hills that are constructed taking into account their lookahead payoff.

5. Data Analysis: Learning by Experienced Players

In this section we turn to an analysis of the experienced players. The basic question we want to address is: what have they learned? In what sense is their behavior qualitatively different from the inexperienced players (if any)? Before considering the specifics of the sessions with experienced players, a relevant question is which players came back.¹³ In other words, how far had the players discovered whether the game was worthwhile?

5.1 Observation 1

The group of players that came back has a selection bias. The better their average profit as inexperienced player, the more likely they were to come back. None of the four bankrupt players came back. Survival as an inexperienced player did not guarantee survival as an experienced player (five went bankrupt). But the players who got experienced in the strongly reduced strategy space (999 instead of 4999 as maximum values for signaling and production) were at a disadvantage.

In Nagel and Vriend (1998) we distinguished three categories of players: the successful players (category I), the ‘nil’ players (category II) and the

¹² There is also an indirect effect related to a change in the player’s sales level. It will change the number of ‘free’ consumers for which the player’s signals compete with the other players’ signals. This indirect effect will be relatively small because it is spread over the six firms (they compete for the same pool of free consumers), and will be ignored here.

¹³ We tried to get all players back. First, they got a letter inviting them to participate in a similar experiment, and a couple of days later we contacted them all by phone.

TABLE 4. Which Players Came Back

Players	Absolute frequencies		Relative frequency (%)
	Yes	No	
All	30	48	38
Category I	22	26	46
Category II	5	13	28
Category III	3	9	25

unsuccessful players (category III). The so-called ‘nil’ players (category II) had both signaling and profit levels close to zero. The successful and the unsuccessful players (categories I and III) had much higher signaling levels, but for the unsuccessful players this came with large losses, whereas the successful players realized positive profits. Table 4 shows the percentages of players that came back, and those that did not, per category. We use a χ^2 -test to compare the probability of category I players coming back with this probability for category II and III players combined. Category I players appear much more likely to come back, but due to the small number of observations this is significant only at the 9.3% level. This might be an example of reinforcement learning. They tried the game as an inexperienced player, and the higher the payoff realized, the more likely they are to play again.¹⁴

Just as with the inexperienced players, with the experienced players too we can distinguish three categories of players: the successful ones, the ‘nil players’ with both signaling and profits close to zero, and the unsuccessful ones. With the inexperienced players we had one treatment in which there was a strongly reduced strategy space: 1000 instead of 5000. Four of the players of those two sessions came back. As inexperienced players these four would have been all four successful category I players, but as experienced players all four of them ended up in category III of the unsuccessful players, and two of them even went bankrupt. This suggests that the other players have learned something in the larger strategy space that turned out useful as experienced players.

Recall that the only difference in the setup with the experienced players is that we changed the parameter values (see Table 1 above). Everything else, including the instructions, remained the same. We explained this to the players. Hence, we had a game with the same structure, but with different numerical values. The reason was that we are interested in the question whether the players had learned something of the general structure of the game, and not whether players that had happened to find good actions as

¹⁴ Garvin and Kagel (1994) also find that aggressive, unsuccessful players are less likely to play again.

inexperienced players would simply continue to choose those specific numbers. A first question, then, is whether the players understood right from the start that their opportunities were different in a quantitative sense.

5.2 Observation 2

In the first period the experienced players' choices for production and signaling levels are less concentrated on focal points than those of the inexperienced players. But that is not because the players simply continue with those values they had learned during the game as inexperienced players.

The production level chosen in the very first period is again rarely greater than 1000; only two players (7%) against one inexperienced (3%). Multiples of 50 are again popular (22 players, or 73%). But multiples of 100 are less popular with experienced players (17 players, or 57%) than with inexperienced (23 players, or 77%), a difference that is significant at 10% (χ^2 -test). The most chosen production levels are 100 (four times), 200 (three) and 300 (three). Hence, the midpoint of 500 no longer has focal point character for production levels. This is a significant difference with the inexperienced players at 1.0% (χ^2 -test). With respect to signaling, multiples of 50 and 100 are still chosen often (25 times, or 83%, and 18 times, or 60%). The most chosen signaling levels are 300, 500, 1000 (three times each, or 10%). Concerning the ratio between signaling and production, 16 inexperienced players out of 30 (53%) chose signaling greater than production, while 24 experienced players (80%) chose signaling greater than production. This is significant at 3.0% (χ^2 -test). Thus, they have learned that signaling is relatively important.

The correlation coefficient between the numbers adapted to as inexperienced players (taking for each individual player his average signaling in the last 50 periods), and the values chosen by exactly the same players in the initial period as experienced players is only 0.14 and not significant (t -test). Hence, already in the first period, the experienced players understand that the different parameters imply different opportunities.

Next, we turn to our two-step model of learning direction theory for production, and hill climbing for signaling. Do the experienced players behave differently? And in particular, we want to know whether the players have learned to look ahead more often than before.¹⁵

¹⁵ To avoid selection bias, when comparing with the inexperienced players, we consider here only those 30 inexperienced players that came back.

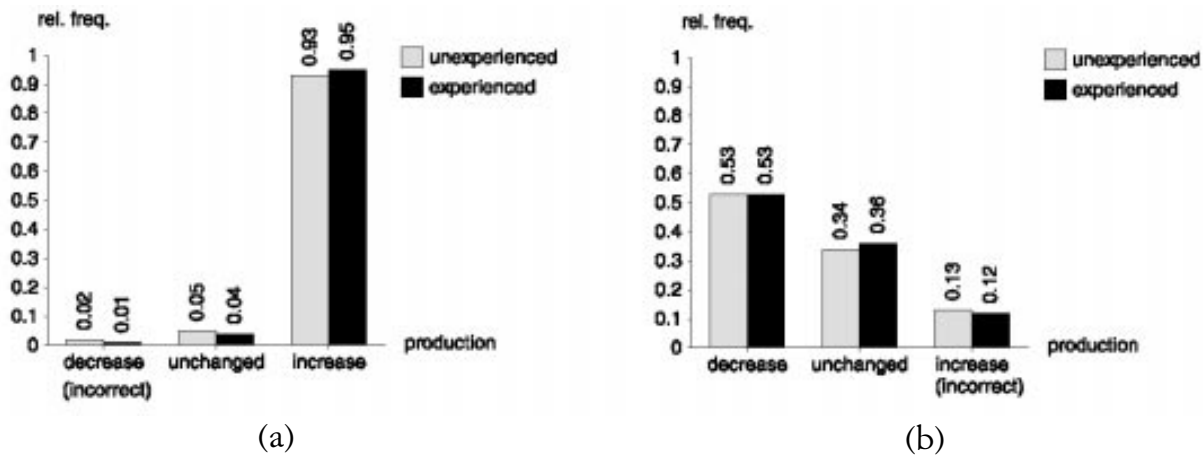


FIGURE 1. (a) Direction learning after production < demand. (b) Direction learning after production > demand.

5.3 Observation 3

The experienced players’ adaptive behavior as described by our two-step model is similar to that of the inexperienced players.

Figure 1(a,b) summarizes how far learning direction theory predicts correctly, distinguishing the cases of too high and too low production in the preceding period, and distinguishing the inexperienced and the experienced players. As we see, with respect to the learning direction hypothesis concerning the adjustment of the production level, there is little difference between inexperienced and experienced players. Inexperienced players adjusted their production level in the wrong direction in 11.4% of the cases, and experienced players in 9.5%. Using the Wilcoxon–Mann–Whitney test (hereafter called the Wilcoxon test) to analyze whether these levels of the individual players are drawn from the same distribution for inexperienced and experienced players, we find that the null hypothesis ‘no difference’ cannot be rejected. As we explained above, *ex ante* optimization would lead to overproduction, which would imply a bias with respect to the learning direction theory hypothesis in the cases where production in the previous period was greater than demand. This bias shows up for both the inexperienced and experienced players. Hence, it seems that they understand equally well the desirability of overproduction.

Figure 2(a,b) shows that the frequencies with which the experienced players tend to climb hills are similar to those for the inexperienced players. Experienced players go in a completely wrong direction with signaling in 28.5% of the cases, compared with 24.2% for the inexperienced for lookahead hill climbing. Using the Wilcoxon test as above, this difference is not significant. As we explained above, myopic players would forget the value of building up their sales level, and they might climb the hill of their immediate

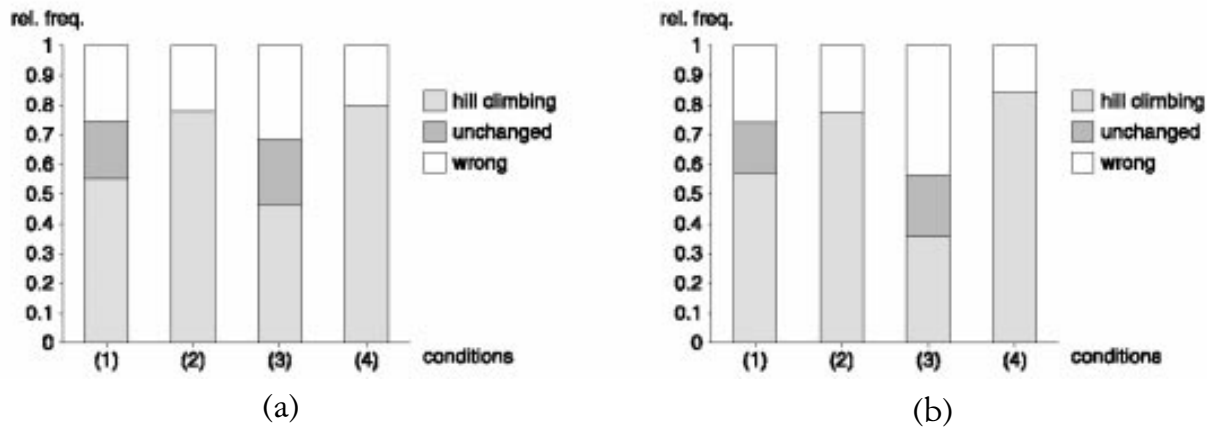


FIGURE 2. (a) Hill climbing inexperienced, with conditions as explained in Table 3. (b) Hill climbing experienced, with conditions as explained in Table 3.

profits only. Analyzing all cases in which a player had changed his signaling level, it turns out that in $\sim 70\%$ of the cases the payoff gradient happens to be in the same direction for myopic and lookahead hill climbing. That is, in those cases the player’s immediate profits as well as his lookahead payoff (taking into account also the future profits related to his current sales level) had increased, or both had decreased. When we consider only the cases of opposite gradients—the cases in which myopic hill climbing and lookahead hill climbing predict a different change in signaling—we find that on average both the experienced and the inexperienced players are inclined only a little bit towards looking ahead, and there is also no significant difference between the inexperienced and experienced players as far as their inclination to look ahead is concerned (Wilcoxon test): 54.3% for inexperienced against 54.2% for experienced players.

In Nagel and Vriend (1998), we showed that there were strong differences between the inexperienced individual players. Since the experienced players came from different sessions, the following two questions arise. Did the experienced players have a significantly different background? And if so, does this imply differences for their behavior as experienced players?

5.4 Observation 4

There were differences in the players’ experience as far as their environment was concerned. And players that got their experience in a more cooperative environment tend to play more cooperatively as experienced players.

Table 5 gives the average signaling per period per player for each of the 11 sessions with inexperienced players. Clearly, in some sessions there was more aggregate signaling than in other sessions. If players signal very little, not all potential customers might be reached, and some might not show up. But very

TABLE 5. Session Averages: Inexperienced

Session	Production	Signaling	Sales	Profits
1	149	690	117	25
2	144	886	116	9
3	143	1028	117	-1
4	157	846	116	9
5	140	1016	116	-1
6	150	909	116	6
7	135	1035	115	-1
8	146	726	116	22
9	135	1043	117	0
10	136	940	117	8
11	139	1010	117	1

TABLE 6. Influence of Environment on Cooperation

Session experienced	Average signaling experienced	From sessions inexperienced	Weighted average signaling inexperienced
21	592	4 from no. 1, 2 from no. 2	755
23	770	2 from no. 3, 1 from no. 4, 3 from no. 9	1005
24	666	2 from no. 5, 4 from no. 6	945
25	769	3 from no. 7, 3 from no. 11	1023

soon almost all customers are reached. If firms were cooperative, they could keep signaling low, but each individual firm would have an incentive to increase its signaling. Hence, we can see the average signaling level in a session as a measure of cooperation among the firms in that session. What is the effect for the individual firms of this level of average signaling? Whatever their sales level, an increase in the aggregate signaling implies that the signaling costs per unit of sales go up. There are two causes for this. First, a firm's sales go down. Second, when a firm reacts, increasing its signaling, its signaling costs go up.

The question, then, is: does this different background imply differences for their behavior as experienced players? Table 6 shows the average signaling per firm for the sessions with experienced players, the numbers of the sessions where they had got their experience, and a weighted average of the average signaling activity in those sessions.¹⁶ The correlation coefficient between the

¹⁶ We omit here session 22 because four players came from sessions in which production and signaling had an upper limit of 999 instead of 4999. That is, they had got their experience in a different environment anyway, independent of their own attitude towards aggressiveness or cooperation.

measure of cooperation as inexperienced and as experienced is 0.94 (significant at 2.9%; one-sided t -test).

6. Conclusions

The main objective of this study was to analyze whether players really learn while playing an oligopolistic market game in a laboratory for many periods. Therefore, we organized two series of experiments—one with inexperienced players and one with experienced players—and analyzed whether there were systematic differences in the behavior between these two groups.

The most surprising result is that the quantitative measures of the two-step model of adaptive behavior are not significantly different between experienced and inexperienced players. In other words, the players learn in the sense that they adapt to their environment, but they do *not* appear to learn in the sense of finding a better method to adapt to their market opportunities. This suggests that the players carry with them an enormous amount of real life experience in a large world, and the extra experience of one experiment means only very little. Another possibility is that the inexperienced players actually do learn to adjust the way in which they adapt to their environment, but that they do not succeed in transferring the skill thus acquired to their new environment when they return as experienced players. The cause of this difficulty is that a player's environment consists in part of the strategies played by the other players. Hence, an experienced player can face a different environment as his new competitors may employ different kinds of strategies, with the result that the player needs to start learning from scratch.

Nevertheless, we did observe some differences between inexperienced and experienced players. We showed that the group of players that came back as experienced players had a clear selection bias, with the successful players being almost twice as likely to play again. Hence, the players had learned whether the game was 'worth playing'. The experienced players' initial period actions had a different distribution to those of the inexperienced players, which suggests that they had learned something, and did not simply start with those values they had adapted to as inexperienced players. And although with respect to our two-step model the experienced players' behavior was not significantly different from that of the inexperienced, we showed that competitiveness of the environment in which the players got their experience had a significant influence on their behavior as experienced players.

Acknowledgements

We wish to thank Reinhard Selten for encouraging and discussing the experiments, and Klaus Abbink, Joachim Buchta, Cornelia Holthausen, Barbara Mathauschek, Michael Mitzkewitz and especially Abdolkarim Sadrieh for their indispensable assistance in discussing and organizing the experiments. We are grateful for comments and discussions concerning previous versions of this paper to Antoni Bosch, Art De Vany, Ido Erev, Steven Klepper, John Miller, Greg Pollock, Phil Reny, Al Roth, Arthur Schram, Ulrich Witt, an anonymous referee, and to seminar and conference participants in Pittsburgh, Ames, Long Beach, Barcelona, Augsburg, Trento, Amsterdam, Toulouse, Marseille, London, St Louis and Paris. The experiments were made possible by financial support from the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 303 at the University of Bonn. Stays at the University of Pittsburgh, and the Santa Fe Institute, and financial support through TMR grant ERBFMBICT950277 (N.J.V.) from the Commission of the European Community are also gratefully acknowledged. The usual disclaimer applies.

References

- Barto, A. G., R. S. Sutton and C. W. Anderson (1983), 'Neuronlike Adaptive Elements that Can Solve Difficult Learning Control Problems,' *IEEE Transactions on Systems, Man, and Cybernetics*, 13, 834–846.
- Bloomfield, R. (1994), 'Learning a Mixed Strategy Equilibrium in the Laboratory,' *Journal of Economic Behavior and Organization*, 25, 411–436.
- Bush, R. R. and F. Mosteller (1955), *Stochastic Models for Learning*. Wiley: New York.
- Camerer, C. and T. Ho (1997), 'Experience-weighted Attraction Learning in Games: A Unifying Approach,' Working Paper 1003, Caltech.
- Carlton, D. W. (1989), 'The Theory and the Facts of How Markets Clear: Is Industrial Organization Valuable for Understanding Macroeconomics?,' in R. Schmalensee and R. D. Willig (eds), *Handbook of Industrial Organization*. North-Holland: Amsterdam, Vol. 1, pp. 909–946.
- Cross, J. G. (1983), *A Theory of Adaptive Economic Behavior*. Cambridge University Press: Cambridge.
- De Vany, A. and W. D. Walls (1996), 'Bose–Einstein Dynamics and Adaptive Contracting in Motion Pictures,' *Economic Journal*, November.
- Garvin, S. and J. H. Kagel (1994), 'Learning in Common Value Auctions: Some Initial Observations,' *Journal of Economic Behavior and Organization*, 25, 351–372.
- Holland, J. H., K. J. Holyoak, R. E. Nisbett and P. R. Thagard (1986), *Induction: Processes of Inference, Learning, and Discovery*. MIT Press: Cambridge, MA.
- Keser, C. (1992), *Lecture Notes in Economics and Mathematical Systems 391: Experimental Duopoly Markets with Demand Inertia: Game-playing Experiments and the Strategy Method*. Springer-Verlag: Berlin.
- Kirman, A. P. (1993), 'Learning in Oligopoly: Theory, Simulation, and Experimental Evidence,' in A. P. Kirman and M. Salmon (eds), *Learning and Rationality in Economics*. Blackwell: Oxford.
- Lazear, E. P. (1995), 'Bait and Switch,' *Journal of Political Economy*, 103, 813–830.

Machine Learning (1992), 8, nos. 3/4, Special Issue on Reinforcement Learning.

Merlo, A. and A. Schotter (1994), 'An Experimental Study of Learning in One and Two-Person Games,' Economic Research Reports 94–17, C. V. Starr Center for Applied Economics, New York University.

Nagel, R. (1995), 'Unraveling in Guessing Games: An Experimental Study,' *American Economic Review*, **85**, 1313–1326.

Nagel, R. and N. J. Vriend (1999), 'An Experimental Study of Adaptive Behavior in an Oligopolistic Market Game,' *Journal of Evolutionary Economics*, **1**, 27–65.

Roberts, M. (1995), 'Active Learning: Some Experimental Results,' mimeo.

Roth, A. E. and I. Erev (1995), 'Learning in Extensive-form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term,' *Games and Economic Behavior*, **8**, 164–212.

Selten, R. and R. Stoecker (1986), 'End Behavior in Sequences of Finite Prisoner's Dilemma Supergames. A Learning Theory Approach,' *Journal of Economic Behavior and Organization*, **7**, 47–70.

Stahl, D. O. (1996), 'Boundedly Rational Rule Learning in a Guessing Game,' *Games and Economic Behavior*, **16**, 303–330.

Stigler, G. J. (1961), 'The Economics of Information,' *Journal of Political Economy*, **69**, 213–225.

Vriend, N. J. (1994), 'Artificial Intelligence and Economic Theory,' in E. Hillebrand and J. Stender (eds), *Many-agent Simulation and Artificial Life*. IOS: Amsterdam, pp. 31–47.

Vriend, N. J. (1996), 'A Model of Market-making,' Economics Working Paper no. 184, Universitat Pompeu Fabra, Barcelona.

Vriend, N. J. (1997), 'Will Reasoning Improve Learning?,' *Economics Letters*, **55**, 9–18.

Appendix

The following list contains the English version of the instructions given to the players.

Actors:

- Each of you will be a firm in a market economy.
- The consumers in this economy are simulated by a computer program.

Each day:

In the morning, firms decide:

- Identical firms decide upon a number of units of a perishable consumption good (each firm the same good).
- The production of each unit costs 0.27 point.
- The production decided upon at the beginning of the day is available for sale on that day.
- Experience shows that, in general, the demand faced by an individual firm is below 1000.
- The firms also decide upon a number of information signals to be sent to the population, communicating the fact that they are a firm offering the commodity for sale on that day. Imagine the sending of letters to addresses picked randomly from the telephone book.

- Sending one information signal to an individual agent always costs 0.21 point.
- The price of the commodity is 1 point. The price of the commodity is given, it does not change over time, it is equal for all firms and consumers, and known to all agents.
- It is not possible to enter values greater than 4999 for the number of units to be produced and the number of information signals to be sent. This is due only to technological restrictions, and has no direct economic meaning.

During the day, consumers are ‘shopping’:

- When all firms have decided their actions, consumers will be ‘shopping’. Each day, each consumer wishes to buy exactly one unit of the commodity. Hence, consumers have to find a firm offering the commodity for sale, and such a firm should have at least one unit available at the moment they arrive.
- We give you two considerations concerning the consumers’ actions:
 - a* A consumer that has received an information signal from you knows that you are a firm offering the commodity for sale on that day.
 - b* Consumers who visited you, but arrived too late and found only empty shelves might find your service unreliable. On the other hand, a consumer who succeeded in buying one unit from you might remember the good service.

At the end of the day, each consumer and each firm observes his own market outcomes:

- Consumers return home satisfied or not, i.e. with or without a unit of the commodity.
- All unsold units of the commodity perish.
- Each firm will know the demand that was directed to it during the day, how much it has actually sold (note that it cannot sell more than it has produced at the beginning of the day), and its profits of that day.
- It cannot be excluded that sometimes the market outcomes are such that a firm makes a loss. Each firm faces an upper limit of 2000 points for the total losses it may realize. A firm exceeding this limit will be declared bankrupt, implying that it will be forced to inactivity from then on.
- A firm might have received some information signals sent to random addresses by other firms. These information signals will be listed (senders and numbers of signals), using fictitious names for the sending firms.

Time:

- There is no time limit for your daily decisions. From day 20 on, you will

hear a warning sound when you are using more than 1 min of decision-time.

- The playing-time will be about 2.5 hours.

Payment:

- Each player will be paid according to the total profits realized by his firm.
- Each player gets a 'show-up' fee of DM 25.
- In addition, the payoff will be DM 12.50 for each 1000 profit points realized.
- Note that losses realized will be subtracted from the DM 25.
- Bankrupt players have lost an amount of DM 25, and hence get nothing.

Anonymity:

- A player will never know the actions and outcomes of other players.

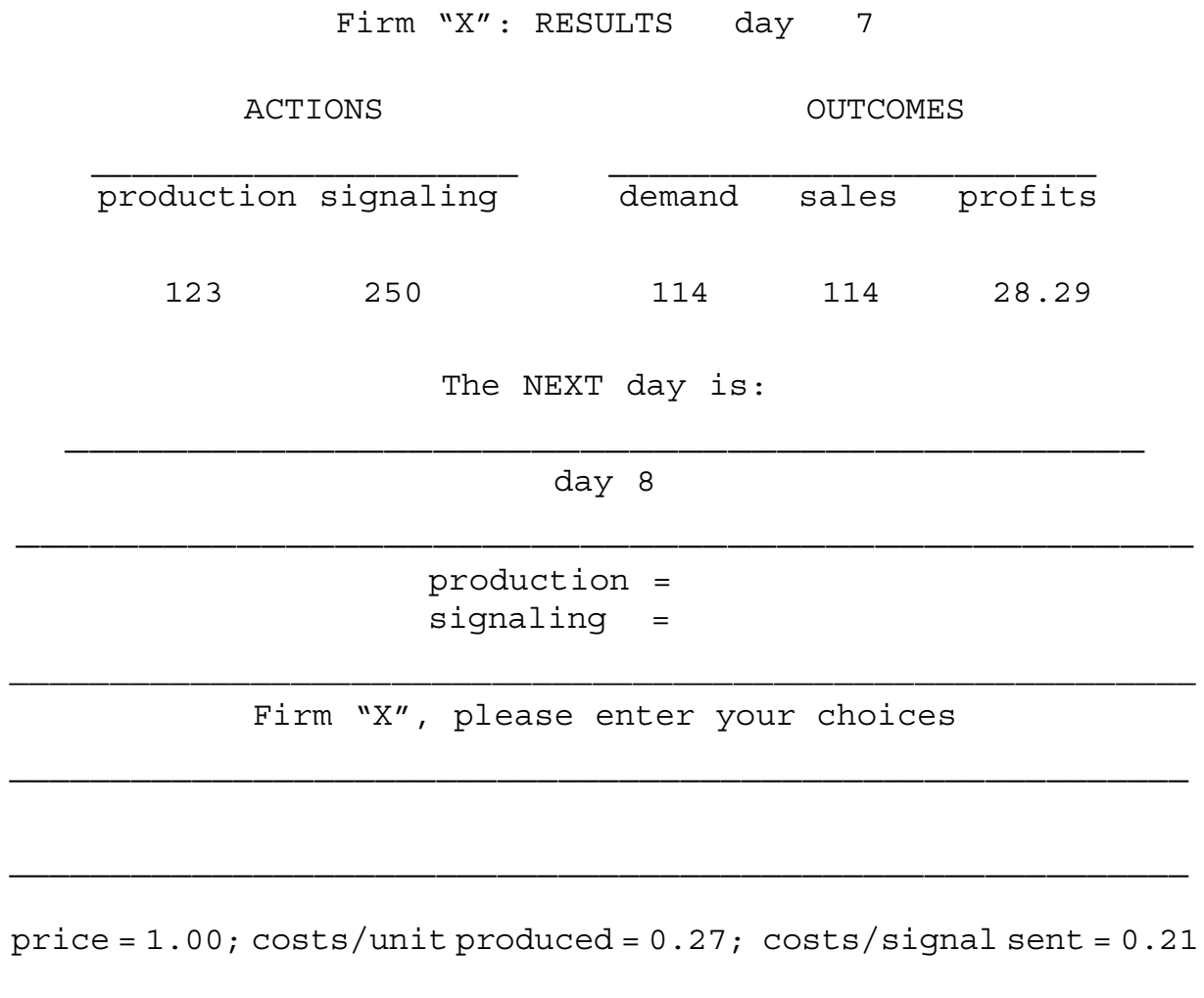


FIGURE A1. Computer screen for firm 'X'.

Keyboard:

- To confirm your choice: Enter [↵]
- To delete: Backspace [←]
- Please, before confirming your choices, always make sure that you did not make a typing-error.

Figure A1 shows the computer screen as viewed by a player acting as firm 'X' in a given period. At the beginning of day 1 the top part of the screen contained the following message: 'Experience (from previous experiments) shows that, in general, the demand faced by an individual firm is below 1000'. When a player had negative cumulative profits, they got a warning in the center of the screen saying: 'WARNING! Your total losses are 192.25 (total losses greater than 2000.00 imply BANKRUPTCY!)'.