An Experimental Test of Equilibrium Dominance in Signaling Games

Jordi Brandts and Charles A. Holt*

Many economic situations with asymmetric information can be modeled as signaling games. Even simple signaling games can have sequential equilibria that are considered "unintuitive". For example, in a well known model due to Michael Spence (1973), workers with good information about their own ability levels have to decide whether to obtain additional education or not. An employer, knowing that education is more costly for low-ability workers, observes the education signal, but not the worker's ability, prior to deciding on a wage offer. As shown below, it is possible to construct examples in which all types of workers decide against the educational investment because the employer will interpret education as a signal of *low* ability, even though the signal is more costly for low-ability workers. Despite the unintuitive nature of these "out-of-equilibrium beliefs", this equilibrium outcome survives the tests imposed by all of the commonly used, strengthened versions of the Nash concept. In particular, this outcome is a sequential equilibrium in the sense of David M. Kreps and Robert Wilson (1982); it satisfies a backwards induction rationality requirement that decisions be optimal from any non-terminal stage until the end of the game, given the equilibrium beliefs at that stage.

Despite the fact that unintuitive outcomes can pass the test imposed by a sequential equilibrium, there is no consensus on exactly how this equilibrium concept should be strengthened or "refined". The recent debate centers on refinements that place more restrictions on players' beliefs about what would happen off of the equilibrium path. Several refinements have been proposed, each of which is motivated by specific games in which a weaker refinement permits unreasonable equilibria. For example, In-Koo Cho and Kreps (1987) discuss their "intuitive criterion", which is based on a dominance notion that they call *equilibrium dominance*. They also discuss a number of stronger refinements, including strategic stability (Elon Kohlberg

---

* Universitat Autonoma de Barcelona and University of Virginia. This research was initiated while Holt was supported by the Comite Conjunto Hispano-Norteamericano sabbatical program. This research was supported by grants from the Spanish DGCICYT (PB86-0613 and PB89-0294) and the U.S. NSF (SES-87-20105). We are grateful to Lisa Anderson, Colin Camerer, Doug Davis, Maxim Engers, and Bill Johnson for helpful comments.

and Jean-Francois Mertens, 1986) and divinity (Jeffrey Banks and Joel Sobel, 1987).[2] We are particularly interested in equilibrium dominance, since it is widely discussed and since it is implied by stronger refinements.

The arguments about which refinement is appropriate are typically based on the author's subjective opinion about how reasonable individuals would behave. There is an apparent need for empirical work that directly tests the validity of these arguments and that could guide the theoretical debate towards alternative refinements. Although game theory is sometimes used as a *normative* theory about how "rational" agents ought to behave, the theory is also widely applied in industrial organization and other areas of economics to analyze specific issues. When game theory is applied in this manner, its *positive*, empirical content becomes relevant. In particular, the discussions of "intuitive" and "unintuitive" equilibria are often in the context of examples that are directly inspired by market applications, e.g. Grossman (1981). Refinements have been subsequently used in the theoretical analysis of a variety of issues: corporate finance (Milton Harris and Artur Raviv, 1985), advertising (Paul Milgrom and John Roberts, 1986), and entry deterrence (Charles Holt and David Scheffman, 1989).

An appropriate laboratory experiment can provide important insights to complement the theoretical debate. In an experiment it is possible to use financial rewards to induce players' preferences in a manner that competing theories give very different predictions about agents' actions.[3] In fact, it is difficult to imagine a situation arising naturally in the economy that would enable an outside observer to distinguish a sequential equilibrium from alternatives of interest.

In this paper we present the results of an experiment that is structured to permit an initial evaluation of equilibrium dominance, a basic intuitive argument that has motivated many of the refinements of the notion of sequentiality. The experiment involves signaling games with two sequential Nash equilibria, one of which is precluded by equilibrium dominance, which is

---

[2] Cho and Kreps conclude that strategic stability and similar refinements are probably too strong in the sense that these theories give predictions (in the context of simple signaling games) that they find to be unintuitive. Sanford J. Grossman and Motty Perry (1986) have developed the concept of a perfect sequential equilibrium, which is neither more or less restrictive than strategic stability, and which may not exist in finite, extensive-form games. Cho and Kreps also discuss the earlier work of John Riley (1979).

[3] There are some recent experimental papers that deal with related questions. The sequential Nash concept organizes the data rather well in multi-stage experimental games with incomplete information reported by Colin Camerer and Keith Weigelt (1988) and Jeffrey Banks, Camerer, and David Porter (1988). The latter authors find support for refinements that are stronger than sequentiality. But in other contexts, i.e. sequential bargaining games, the predictive power of the sequential equilibrium is diminished by concerns for distributional fairness (Jack Ochs and Alvin Roth, 1989).

explained in section I.[4] The experimental procedures are described in section II. Section III contains results of a treatment based on a slight modification of a signaling game that has been widely discussed in the refinements literature. Observed decisions roughly correspond to the intuitive equilibrium for this treatment. A second treatment, which was motivated by an examination of the adjustment process under the first treatment, is reported in section IV. This second treatment involves a basic design in which the subjects' initial behavior can generate a relationship between signals and players' types that corresponds to the out-of-equilibrium beliefs of the unintuitive equilibrium. After gaining experience with different partners in a series of these signaling games, behavior closer to the *unintuitive* equilibrium outcome is observed. As indicated in the final section, we believe that these results constitute a challenge to theorists and a rationale for further study of out-of-equilibrium adjustment processes.

I. The Equilibrium Dominance Criterion

Cho and Kreps (1987) discuss the ideas behind equilibrium dominance in the context of signaling games with two players. In this class of games, one of the players can have one of a number of different preference "types". The second player's payoffs can depend on the first player's type, but this other player does not know the first player's type. Given this asymmetry of information, the first player, who knows his own preference type, sends a message (signal) to the second player, who then takes an action. In these games, one typically finds many Nash equilibria due to the existence of many off-the-equilibrium-path beliefs, i.e. many inferences that the second player could make after observing a message that, in equilibrium, would *not* be sent by the first player. One way to eliminate some of the beliefs is to ask the following questions:

1. Which out-of-equilibrium messages should not "reasonably" be expected when the first player is of a certain type?
2. Which actions are "unreasonable" as responses to certain out-of-equilibrium messages?

By eliminating combinations of unintuitive messages and responses, we can delete some of the Nash equilibria. We begin by considering the issue raised in the first question above, i.e. the elimination of unreasonable out-of-equilibrium messages; the discussion of unreasonable responses to out-of-equilibrium messages is considered at the end of this section.

Messages

To apply the notion of equilibrium dominance, one concentrates attention on a particular

[4] The first experiment pertaining to signaling behavior in markets was conducted by Ross M. Miller and Charles R. Plott (1985).

equilibrium and the corresponding equilibrium payoff that the first player obtains when he is of a particular type. According to equilibrium dominance, one can eliminate a combination of a type and an out-of-equilibrium message if the *equilibrium* payoff for the player of that type is larger than the *highest* payoff he could get if he were to send the out-of-equilibrium message under consideration. The equilibrium under consideration is then ruled out if the beliefs that support it place positive probability on the preference-type/message combination that was eliminated in this manner.

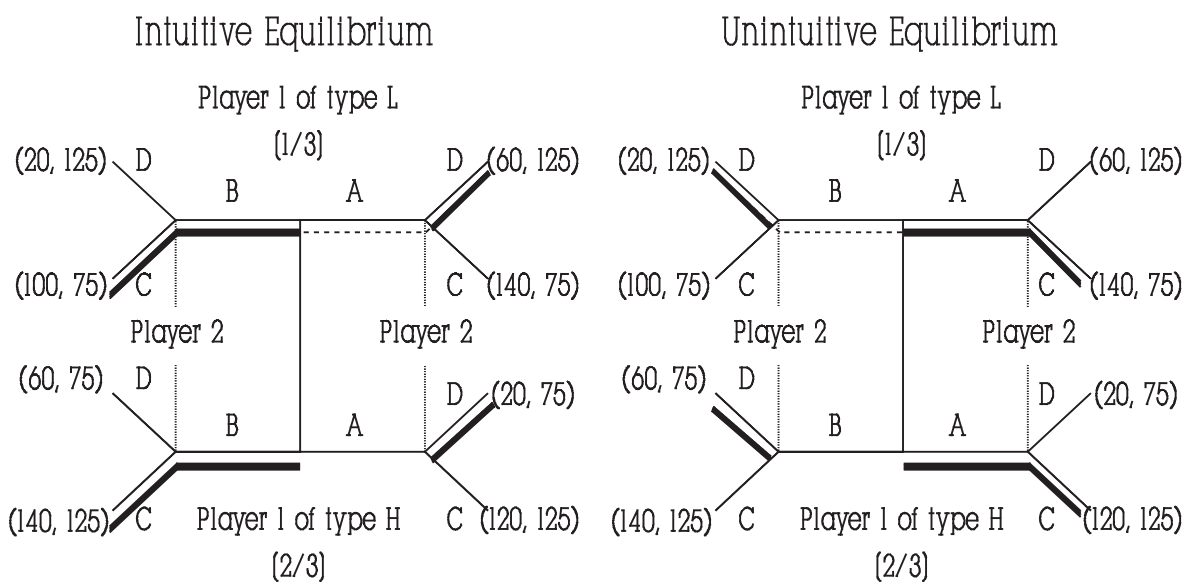

Figure 1. Two Pooling Equilibria for Game 1
Key: boldfaced line represents the equilibrium decisions
dashed line represents out-of-equilibrium path

The notion of equilibrium dominance can be illustrated by considering a signaling game, which will be called game 1, in which this refinement rules out one of the two equilibrium outcomes. The extensive form for this game is reproduced on both sides of figure 1. One of the equilibria is represented on the left side, and the other is represented on the right. The game begins with nature's determination of player 1's type, and the fractions in square brackets indicate that the type is L with probability 1/3 and H with probability 2/3. The messages for player 1 are denoted by A and B, and the responses for player 2 are denoted by C and D. The dotted, vertical lines determine player 2's information sets; he observes the message, A or B, but not player 1's type, H or L. The payoffs are given in parentheses at each terminal node, with player 1's payoff listed first.

In this game it is straightforward to verify that there are only two pure-strategy, sequential equilibrium outcomes. In the equilibrium shown on the right side of the figure, both types of player 1 choose message *A*, and player 2 responds to *A* with *C* and to *B* with *D*. The equilibrium strategies for the two players are indicated by the boldfaced lines. The other equilibrium, shown on the left side of the figure, is one in which both types choose *B* and player 2 responds to *B* with *C* and to *A* with *D*.

First consider the incentive of player 2 to deviate from the decision *C* in either equilibrium. If both types of player 1 choose the same message, then it follows from the symmetry of player 2's payoffs that *C* is the optimal response, since the common message is more likely to be sent by a player 1 of type *H*. The incentive for player 1 to deviate depends on the anticipated response of player 2, which in turn depends on how player 2 interprets the out-of-equilibrium message. Each of the "pooling" equilibrium outcomes under consideration is supported by player 2's beliefs that a deviant message is more likely to come from a player 1 of type *L*, so player 2 would respond to a deviation by choosing *D*, which deters deviations by either type of player 1. The belief that a deviant signal is more likely to come from a type-*L* proponent is represented by the horizontal dashed line. In the equilibrium in which both types send the message *A*, for example, it is apparent from player 2's payoffs that the best response to an out-of-equilibrium message *B* is *D* if the probability is greater than 1/2 that this deviant message was sent by a type *L*. Player 2's beliefs, that the out-of-equilibrium message *B* is more likely to come from a player of type *L*, do not violate the relatively weak consistency-of-beliefs requirements of a sequential equilibrium.

This equilibrium on the right side of figure 1, in which both types of player 1 send message *A* and player 2 responds with *C*, can be ruled out by applying the notion of equilibrium dominance, however. As noted above, equilibrium dominance involves an analysis of out-of-equilibrium beliefs by making a comparison of a player's equilibrium payoff with the best payoff that could be obtained by deviating, so we begin by noting that the payoffs in this equilibrium (with both types choosing message *A*) are 140 for a player 1 of type *L* and 120 for a player 1 of type *H*. Recall that this equilibrium is supported by beliefs that a deviant sending message *B* is more likely to be of type *L*. But if the type-*L* player were to send the out-of-equilibrium message *B*, the most he could possibly obtain would be 100 (resulting from a response of *C*), and this payoff is less than the type *L*'s equilibrium payoff of 140, which results from a message of *A* followed by the equilibrium response of *C*. Thus equilibrium dominance rules out the possibility that type *L* will send the out-of-equilibrium message *B*. Since it can be verified that the out-of-equilibrium message *B* is not equilibrium dominated for the type-*H* player 1, player 2 will reason that message *B* can only come from a player 1 of type *H*. Thus player 2 would react to *B* with *C* in order to obtain earnings of 125 instead of 75.

But then a player 1 of type *H* will have an incentive to deviate by sending message *B*, since this deviation would raise this player's payoff from 120 to 140. This breaks the equilibrium.

The other equilibrium outcome, in which both types of player 1 send message *B*, survives such a test because the out-of-equilibrium message *A* is not equilibrium dominated for the type-*L* player 1 Therefore, player 2's beliefs, that a deviation is more likely to come from a type *L*, are not unreasonable in this equilibrium.

We intentionally discussed the above example in a neutral, non-economic context, so as to stress the role of the relative payoffs and probabilities in eliminating one of the equilibria. However, the reader may wish to consider the labor-market interpretation that was alluded to in the introduction: The worker either has a high skill level (type *H*) or a low skill level (type *L*), and knowing this, must decide whether to make an investment in education (signal *B*) or not (signal *A*). Player 2 is an employer who observes the signal but not the worker's type. The employer's payoffs in figure 1 indicate that his objective is to match a skilled worker with the executive job (*C*) and to match an unskilled worker with the manual job (*D*); education is not productive since the employer's payoffs are unaffected by the signal. Both worker types would prefer the executive job *C* for any given level of education. For a given job assignment, the skilled, type-*H* worker obtains a higher payoff with an education, and the unskilled worker obtains a higher payoff without an education, so education is more costly for the type-*L* worker. There is no sequential "separating" equilibrium in which the skilled worker obtains an education and the unskilled worker does not. This is because the unskilled workers would deviate and obtain an education in order to obtain the job *C* that would be assigned to the skilled worker if there were a separating equilibrium. In the intuitive pooling equilibrium, both worker types obtain an education, and out-of-equilibrium beliefs that support this equilibrium are that anyone who does not obtain an education is more likely to be unskilled. But in the unintuitive pooling equilibrium, neither worker type obtains an education, and the out-of-equilibrium beliefs are that a deviant with an education is *more* likely to be an unskilled worker.

It is straightforward to show that the unintuitive equilibrium in this example cannot be ruled out with the weaker notion of simple dominance.[5,6] But this equilibrium can be eliminated

---

[5] It is apparent from figure 1 that (simple) dominance would not eliminate the unintuitive Nash equilibrium in which both types send message *A*, because the *lowest* payoff that player 1 can get when he is of type L and sends the equilibrium message *A* is 60, which is not greater than the *highest* payoff, 100, that could be obtained by sending the out-of-equilibrium message *B*. Equilibrium dominance arguments are less intuitive than simple dominance arguments.

[6] This game can also be used to illustrate the difference between equilibrium dominance in the extensive-form representation and iterated weak dominance in the normal-form representation. The relevant normal form is for a 3-person game since player 1, knowing his own type, corresponds to two players in the game with incomplete information. This is because player 1 cannot commit himself to a strategy (a decision to be made by each of his types) before learning

with theories that generate refinements that are stronger than the refinement resulting from the application of equilibrium dominance. For example, it is apparent from the discussion in Cho and Kreps that both the Banks and Sobel notion of divinity and the Kohlberg and Mertens notion of strategic stability will also rule out the unintuitive equilibrium in this example. Thus experiments based on the signaling game represented in figure 1 are relevant for the evaluation of these stronger refinements as well.

Responses

We now consider the elimination of unreasonable responses to out-of-equilibrium messages; the issue mentioned in the second question at the beginning of this section. In particular, a dominated response is unlikely to be observed. For example, consider game 1' that is constructed from game 1 in figure 1 by adding a third response *I* (for ***I***rrelevant) to each of player 2's information sets. The payoffs determined by this response are the same as if player 2 had selected *C* and given his payoff to player 1. For example, if a type-*L* player 1 were to make a decision of *B* and receive a response of *C*, the payoffs would be (100,75), as indicated in the appropriate node of figure 1. The payoffs for the *I* response to the signal *B* would be (175,0), i.e. player 2 "gives" his payoff to the other player. Similarly, the *I* response to a *B* decision made by a type-*H* player 1 yields payoffs of (265,0), etc. Since player 2 is sure to obtain a positive payoff with either of the other decisions, the *I* decision is dominated and should not be observed.

Once the "give-away" decision is removed from the game, the analysis is the same as before; the (*A*,*C*) equilibrium is broken by an equilibrium dominance argument. But if the *I* decision is not first pruned, it is no longer the case that a deviation of *B* from the (*A*,*C*) equilibrium is equilibrium dominated for the player 1 of type *L*. This is because the equilibrium payoff for the type-*L* player 1 in the (*A*,*C*) equilibrium is 140, but this is no longer greater than the highest possible payoff that this player could obtain by deviating when the give-away decision is included for player 2. Cho and Kreps suggest that the dominated response be pruned before the equilibrium dominance criterion is applied, in accord with question 2 in section 2 above.

---

which type is relevant. It is straightforward but tedious to verify that iterated weak dominance cannot eliminate the unintuitive Nash outcome in the resulting 3-person game in normal form. If the structure of the game permitted player 1 to commit himself to a type-contingent strategy, then iterated weak dominance could be used to eliminate the unintuitive outcome from the resulting normal form of the relevant 2-person game.

II. Procedures and Experimental Design

The signaling game shown in figure 1 was used in the first treatment to be discussed. The payoffs were made in either Spanish pesetas or U.S. pennies, as noted below. One U.S. dollar was worth between 110 and 130 pesetas at the time of the experiment (1987-8).[7] Groups of subjects were recruited for two-hour sessions from undergraduate economics classes, either at the Universitat Autonoma de Barcelona (sessions 1-5), or at the University of Virginia (sessions 6-9).[8] Subjects were told that they would receive an initial payment (1000 pesetas in Barcelona, $5.00 in Virginia) in addition to all cash earnings obtained during the two-hour session. The session began with a reading of the instructions in the appropriate language; a copy of the instructions is available from the authors on request.

In comparison with most other laboratory experiments, ours was not unusually complex, and the subjects did not appear to be confused by the instructions. Their questions were simple and easily answered by paraphrasing the relevant part of the instructions. Each two-hour session involved either 8 or 12 participants, as noted below, and one "monitor", who was one of the subjects selected for this task by the throw of dice at the beginning of the session. The monitor's role was to throw the die that determined the preference types and to observe and ensure that the session was being conducted in accordance with the procedures specified in the instructions. The subjects were placed in two adjoining rooms, with half of them in one room having the role of player 1, the "proponent", and the other half having the role of player 2, the "respondent".

The experimental session consisted of a series of matchings, with each subject being paired with a different partner in the other room in each matching. At the beginning of each matching the monitor would throw a die in the proponents' room. The numbers 1 and 2 determined type L for the proponents, and the numbers 3-6 determined type H. Each proponent then recorded his message, A or B, on the record sheet provided, and this message was privately communicated to the paired respondent in the other room by writing the proponent's message on the respondent's record sheet. Then each respondent would mark a

---

[7] The payoffs in figure 1 were computed by applying an affine transformation to the payoffs in figure 1 in Cho and Kreps (1987), with the exception that player 1's type-H payoff for the (A,C) outcome in our figure 1 is 20% greater than the payoff obtained from the transformation. This was done to raise player 1's expected payoff for the "unintuitive" (A,C) equilibrium to the level of the expected payoff for that player in the other (B,C) equilibrium. Of course, we had no reason to believe that subjects were risk neutral. But the previous section's analysis of these two equilibria is independent of assumptions about risk preferences. In particular, the symmetry of the respondent's payoffs implies that the best response is C whenever the probability that the observed signal was sent by a type-H proponent exceeds .5 , as is the case in either pooling equilibrium.

[8] Two of the 80 participants were economics graduate students. Few, if any, of the subjects had been exposed to elementary game theory.

decision, *C* or *D*, on his or her record sheet, with the response being communicated to the appropriate proponent in the other room. Finally, each respondent was informed of the type of proponent he or she was paired with, and all subjects calculated their earnings with the tables provided in the instructions.[9]

Each subject's record sheet indicated the identification number of the subject in the other room with whom he or she was paired in that matching, and hence, they could see that they would deal with a different partner in each matching. Subjects were not given any information about the rotation schedules of other subjects. This was done to preserve the one-period nature of the games.[10] After four matchings, each subject had been paired with each of the subjects in the other room, thereby completing the first part, "part a", of the session. Subjects were told initially that the first part would be followed by a different game, with different subject identification numbers, and the instructions for a second part, part b, were not distributed until the termination of part a. The game used for part b was game 1' with the third, dominated response, *I*, as discussed in the previous section. In part b, the subjects who had been respondents in part a were given the role of proponents, and vice versa, and the nature of the treatment was disguised by altering the labeling and order of the decisions in the payoff table and by adding 15 pesetas to each payoff. After four matchings of part b, the instructions for a different game to be played in the four matchings of part c were distributed, again with new identification numbers. Part c was the same signaling game that was used in part a, with the roles of proponent and respondent reversed.[11]

---

[9] Subjects were not asked to write explanations of their decisions, except in treatments 3-5, as discussed in footnote 18 below.

[10] With rotation information, a proponent could determine whether dealings with the respondent in the current matching could ever have an indirect effect on the behavior of a (different) respondent to be encountered in a later matching. Proponent number *X*, for example, could determine whether the respondent with which he or she is paired in the first matching will meet a particular proponent in the second matching who will meet the specific respondent in the third matching who is paired with proponent *X* in the fourth matching. We believe that ignorance of rotation schedules is sufficient to prevent the possibility of such a cycle from affecting the proponent's first-matching message. The cycle requires 4 matchings, so the decisions in the final 3 matchings would not be affected by the possibility of such a feedback effect.

[11] We believed that role reversal would be important, but we did not do the direct reversal of roles in part a, with no other change in the game structure, until part c. The instructions promised a new treatment after part a, so we added the third response and changed the labeling of decisions to minimize the similarity between parts a and b.

Table 1. Experimental Design

	treatment 1 sessions 1-4	treatment 2 session 5
part a (matchings 1-4)	game 1	game 1 with announcements
part b (matchings 5-8)	game 1' (roles reversed)	game * (roles reversed)
part c (matchings 9-12)	game 1 (roles as in b)	game 1 with announcements (roles as in b)

treatment 3 session 6	treatment 4 session 7	treatment 5 sessions 8-9
game 2N	game 3N	game 3R: matchings 1-6
game 2R (roles reversed)	game 3R (roles reversed)	________________
game 2R (roles reversed)	game 3R (roles rversed)	game 3R: matchings 7-12 (roles reversed)

The structure of the experiment is summarized in table 1. As indicated in the table, the treatment 1 that was explained above was used with four 8-subject cohorts, sessions 1-4. The other treatments, to be discussed in later sections, are listed across the top row. The games that correspond to parts a-c are listed in the column below the treatment number. The main structural change occurs in treatment 5, which only has two parts, one for matchings 1-6 and another for matchings 7-12. Role reversals are indicated in parenthetical remarks.

III. Results for Treatments 1 and 2

Table 2 presents the type-contingent signals and the signal-contingent responses for each matching of treatment 1. The row for matching 1 of part a indicates that there were 8 proponents of type *H*, all of whom sent signal *B*, and 8 proponents of type *L*, only one of whom sent signal *B*.[12] Signal *B* was always followed with the *C* response, and signal *A* was followed with the *D* response for 6 out of 7 pairs of subjects in the first matching. In the intuitive equilibrium that is not ruled out by equilibrium dominance, all proponents of each type send the *B* signal, but we observe that, in this first matching of part a, all 8 type-*H* proponents sent signal *B*. Therefore all *A* signals were sent by type-*L* proponents, as was the case in subsequent matchings. This type dependence is consistent with the out-of-equilibrium beliefs that support the intuitive sequential equilibrium; i.e. beliefs that a deviant signal, *A*, is more likely to be sent by a type-*L* proponent.

It is interesting to consider how this type dependence might arise. Notice that if the proponent (player 1) has type-*L* preferences in figure 1, then decision *A* is better than *B* if the respondent chooses *D*, and decision *A* is also better than *B* if the respondent chooses *C*. This suggests that a type-*L*'s deviation to decision *A* might be motivated by the belief that the "signal", *A* or *B*, will have no effect on the respondent's decision. These beliefs are contradicted by the actual decisions of the respondents, since the *B* signal is always followed by a *C* response, and the *A* signal is followed by a *D* response in 6 of the 7 cases. Recall that, in the intuitive (*B*,*C*) equilibrium, a deviation to *A* will generate a *D* response.

Reading down the rows for subsequent matchings in table 2, we see that type dependence tends to diminish with experience as type-*H* proponents continue to send signal *B* and type-*L* proponents begin to switch away from signal *A*. As can be seen from the summary percentages at the bottom of the *B*/*L* column, the proportion of type-*L* proponents that send *B* rises from 21% in part a to 58% in part b and 75% in part c. This is consistent with the observed responses to signals, a risk-neutral subject would prefer to send signal *B* if his beliefs matched the

[12] Since there are 4 sessions and 4 pairings in each matching of a session, it follows that there are 16 proponents' decisions for each matching in part a. In part b, there are only 12 observed outcomes for each matching because the part b of one of the sessions used a different game (not reported here), which did not involve signaling. There are 16 outcomes for game 1 in each matching of part c.

Table 2. Type-Contingent Signals and Signal-Contingent Responses for Treatment 1

	signals given type:		responses given signal:	
part/matching	*B* / *H*	*B* / *L*	*C* / *B*	*D* / *A*
a / 1	8 / 8	1 / 8	9 / 9	6 / 7
a / 2	8 / 8	2 / 8	9 / 10	2 / 6
a / 3	12 / 12	1 / 4	13 / 13	3 / 3
a / 4	12 / 12	1 / 4	13 / 13	3 / 3
b / 5	12 / 12	-	12 / 12	-
b / 6	4 / 4	4 / 8	8 / 8	4 / 4
b / 7	12 / 12	-	12 / 12	-
b / 8	8 / 8	3 / 4	11 / 11	1 / 1
c / 9	16 / 16	-	16 / 16	-
c / 10	12 / 12	3 / 4	15 / 15	0 / 1
c / 11	4 / 4	9 / 12	12 / 13	2 / 3
c / 12	12 / 12	3 / 4	15 / 15	1 / 1
percent for part a	100 percent	21 percent	98 percent	74 percent
percent for part b	100 percent	58 percent	100 percent	100 percent
percent for part c	100 percent	75 percent	98 percent	60 percent

observed aggregate distribution of responses, i.e. the proportions of *C* given *B* and of *D* given *A* in table 2.

Table 3. Data for Treatments 1 and 2 by Part

Treatment 1:

	part a game 1			part b game 1’				part c game 1	
	C	*D*		*C*	*D*	*I*		*C*	*D*
A	5 (0,5)	14 (0,14)	*A*	0	5 (0,5)	0	*A*	2 (0,2)	3 (0,3)
B	44 (39,5)	1 (1, 0)	*B*	43 (36,7)	0	0	*B*	58 (44,14)	1 (0,1)

Treatment 2:

	part a game 1 (with announcements)		part b (game *, not reported)		part c game 1 (with announcements)	
	C	*D*			*C*	*D*
A	6 (2,4)	5 (2,3)		*A*	3 (0,3)	6 (1,5)
B	3 (3,0)	2 (1,1)		*B*	5 (2,3)	2 (1,1)

Key: N
(N_H, N_L) indicates N outcomes, N_H with type-*H* proponents and N_L with type-*L* proponents.

In order to facilitate comparisons with the results of subsequent treatments, the outcomes for each part of treatment 1 are given at the top of table 3. Each entry in the table consists of the total number of outcomes for the pair of decisions determined by the row and column, with the breakdown by preference type given in parentheses: (outcomes in type-*H* matchings, outcomes in type-L matchings). Only 7 decision pairs in treatment 1 matched the unintuitive (A,C) equilibrium, and all but 2 of these occurred in the first part with inexperienced subjects. The 112 games conducted in parts b and c with experienced subjects provide no support for the unintuitive equilibrium; 101 of the decision pairs match the other (B,C) equilibrium, and all but 2 of the proponents' deviations were followed by the D response that is consistent with the beliefs off the equilibrium path that support this other equilibrium.[13]

Having failed to generate a preponderance of (A,C) outcomes in treatment 1, we decided to modify the instructions by adding a suggestion that players use the strategies specified by the unintuitive equilibrium for the game being used. The setup is summarized in the column for treatment 2 in table 1; game 1 is used in parts a and c.[14] In part b, we used a different game, game *, which is discussed in detail in Brandts and Holt (1989) and will not be reported here.[15] In each of the three parts, the monitor was asked to read an announcement suggesting that proponents send signal A and that respondents answer A with C and B with D. The announcement ended with a statement that is translated: "This proposal is only a suggestion, not

---

[13] The deviations cannot be interpreted as the result of the behavior of one or two subjects who were systematically using the unintuitive equilibrium strategy. There were 5 deviations in part c of the sessions using treatment 1; each was made by a different subject.

[14] In this session and in all subsequent ones to be discussed, the monitor threw the die individually for each proponent. We thought that this would reduce variability. The equilibrium calculations are not affected.

[15] Game * was designed to evaluate the possibility that the results of treatment 1 were not due to equilibrium dominance considerations, but rather, are due to salient aspects of the payoff structure. For example, although the expected payoffs for the proponent as well as for the respondent are the same in each equilibrium in game 1, the (B,C) outcome maximizes the joint payoff in the most likely state (type H). Therefore we constructed game * in such a way that (A,C) maximizes the joint payoff in either state. In this game, which was called "game 2'" in Brandts and Holt (1989), the respondent's payoffs that follow signal B in game 1 were reduced by 40 pesetas, and the respondents' payoffs that follow signal A were increased by 25. This gives the respondent a strong preference for receiving signal A, since the respondent's lowest payoff after A is larger than his highest after B. In addition, another minor change was made to restore a superficial symmetry that was present in the original Cho and Kreps example. These changes do not alter the difference between the payoffs that the respondent can get for a given signal, nor do they alter the equilibrium calculations and equilibrium dominance arguments. We ran one session with game * (with no announcements of suggested play), both with and without the addition of a third, dominated response. In comparison with treatment 1, there were more D responses to the respondent's least preferred (B) signal, but such "punishments" were not given by experienced respondents in part c, where the (B,C) outcome is prevalent. Similar results were obtained in subsequent session (also reported in Brandts and Holt, 1989) in which the respondent's payoffs were adjusted to make it less costly to punish signal B if the proponent turns out to be of type H. When game * was run in part b of treatment 2, with an announcement suggesting the unintuitive equilibrium, the outcomes were equally divided between A and B signals, with all C responses.

a requirement".[16] The monitor read the announcement in both rooms and told the subjects that the same announcement was read in both rooms.[17] A single session was conducted under this treatment 2, and the data are shown in the bottom part of table 3. The most common outcome in part a is the (*A*,*C*) outcome, but the relative frequencies of the (*A*,*C*) and (*B*,*C*) outcomes are reversed between parts a and c. Moreover, the *A* signal was met with the *D* response in six of the nine cases in part c. We believe that this experience would have resulted in fewer *A* signals if the session had continued with more matchings.

IV. Reverse Type Dependence and "Unintuitive" Results

Behavior in early matchings of treatments 1 and 2 typically does not correspond to any of the equilibria. The experience gained in subsequent matchings with different players leads to a process of adjustment in behavior and beliefs. In order to understand this process, it is necessary to look at the evolution of decisions.

Table 2 reveals a pattern of adjustment for treatment 1, which was also observed for game * discussed in footnote 14: in the early matchings, the type-*L* proponents usually chose *A* and the type-*H* proponents almost always chose *B*. It is apparent from the first row of table 2 that respondents seemed to anticipate this type dependence even in the first matching; they responded to *B* with *C* and to *A* with *D*. This "punishment" for those who sent the *A* signal caused the type-*L* proponents to switch to the *B* signal in later matchings, which led to the intuitive pooling equilibrium. The labor-market interpretation of this type dependence is that only high-ability workers obtain an education in early matchings, and knowing this, the employer assigns a worker with no education to the manual job. In order to obtain the preferred job, the low-ability workers would want to invest in the education signal.

An obvious question is what would happen in a session in which we could somehow induce a "reverse" type dependence in early matchings, with the type-*L* proponents choosing signal *B* and the type-*H* proponents choosing signal *A*. In the labor-market interpretation, this

---

[16] A translation of the main part of the announcement is: "In my role as monitor, I will read this announcement in this room and in the room with the other participants. I have been asked to make the following proposal with respect to the options to be chosen by the participants in the experiment. It is proposed that the proponent always choose *A*, regardless of the earnings table which has been selected by the throw of the die. After communicating the choice of *A* to the respondent with whom the proponent is paired, the respondent will choose, according to this proposal, the option *C*. However, the respondent will choose *D* if the proponent does not choose *A*. In summary, ...Although this proposal will be communicated to all participants, all of you may take the decisions that you think to be appropriate. This proposal is just a suggestion, not a requirement."

[17] John B. Van Huyck, Anne B. Gillette, and Raymond Battalio (1988) and Jordi Brandts and Bentley MacLeod (1989) have recently used similar announcements in experiments.

reversal would mean that only a low-ability worker obtains an education and is, therefore, assigned to the manual job. This gives a low-ability worker the incentive to signal differently in order to be pooled with the uneducated, high-ability workers in executive jobs. In this way, the unintuitive (no-education) equilibrium might be reached after a sequence of matchings. The designs of games used in treatments 3-5 were motivated by these considerations.

First, consider the decision of a risk-neutral proponent of either type. The optimal decision will be a function of the proponent's initial beliefs about the relationship between signals and responses. Let $P(C|A)$ and $P(C|B)$ denote a proponent's prior probabilities that signals of A and B respectively are followed by response C, and let the proponent's payoffs for the four possible combinations of signals and responses be denoted by U_{AC}, U_{AD}, U_{BC}, and U_{BD}, where these payoffs will depend on the proponent's type, H or L. Then the optimal decision for the proponent is to choose signal A if

(1) $P(C|A)U_{AC} + [1 - P(C|A)]U_{AD} > P(C|B)U_{BC} + [1 - P(C|B)]U_{BD}$.

A reasonable conjecture is that one would tend to observe reverse type dependence if there is a "large" set of prior probabilities about respondents' behavior for which the optimal decisions are B for type-L proponents and A for type-H proponents.

Treatment 3

The left side of figure 2 contains the extensive form for game 2*R*, where the "*R*" designation is used to indicate that reverse type dependence is expected (as explained below). In comparison with the baseline game 1 of figure 1, several qualitative changes should be noticed. First, the differences between the respondent's payoffs for the pairs of C and D choices have been increased from 50 in game 1 to 100 in game 2*R*, which gives the respondent a greater incentive to guess the proponent's type correctly. Second, $U_{BD} > U_{AD}$ for a type-L proponent and $U_{BD} < U_{AD}$ for a type-H proponent in game 2*R*, where both inequalities have been reversed relative to the structure of the baseline game 1. By raising U_{BD} for type-L proponents and lowering it for type-H proponents, we have, in a loose sense, attempted to make B more attractive for a type-L proponent and to make A more attractive for a type H, which increases the likelihood of reverse type dependence. Notice, however, that the two pooling equilibria indicated by the boldfaced lines in figure 1 are also equilibria for game 2*R*. In particular, since the proponent is more likely to be of type H, the optimal response is C when both types send signal A, and the supporting beliefs that a deviant signal B is more likely to be sent by a type-L proponent are unintuitive in the sense that the equilibrium payoff of 190 for

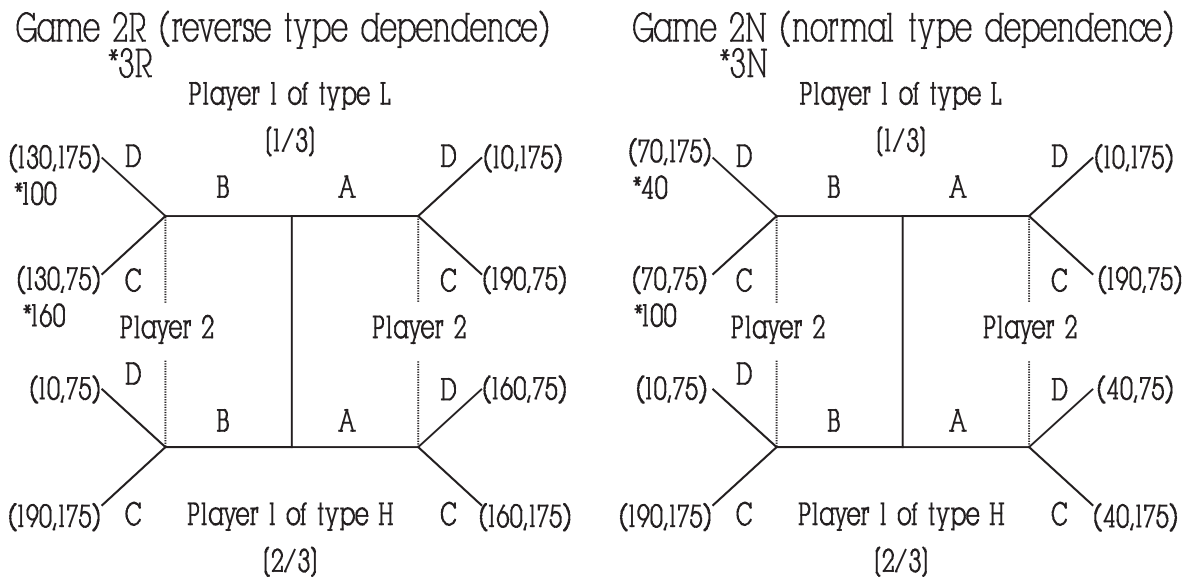

Figure 2. Signaling Games with Normal and Reverse Type Dependence

a type-*L* proponent dominates all payoffs that this player could obtain with a deviant signal, *B*.

The optimal decisions of a risk-neutral proponent for game 2*R* are functions of the prior probabilities, $P(C|A)$ and $P(C|B)$, that are represented on the horizontal and vertical axes of the probability box at the left side of figure 3. Using the type-*L* payoffs from game 2*R* (U_{BD} = 130, U_{BC} = 130, U_{AD} = 10, U_{AC} = 190), one can use inequality (1) to show that signal *A* is optimal if $P(C|A) > 2/3$, which corresponds to the area to the right of the dotted, vertical in the interior of the box on the left side of figure 3. Similarly, a risk-neutral type-*H* proponent prefers choosing *A* if $P(C|B) < 5/6$, as shown by the horizontal dotted line in the box on the left side of figure 3.

The two dotted lines in the interior of the probability box for game 2*R* divide this box into four regions (please ignore the dashed line for now). Both types choose signal *B* in the northwest region, which corresponds to the intuitive equilibrium, and both types choose *A* in the southeast region, which corresponds to the unintuitive equilibrium. There is "reverse" type dependence in the relatively large, southwest region, and there is "normal" type dependence in the smaller, northeast region. If reverse type dependence were observed by respondents in early matchings, they would be likely to respond with *D* to a *B* signal and with *C* to an *A* signal, which could cause the type-*L* proponents to switch from *B* signals to *A* signals. In this sense, reverse type dependence in the southwest region of the probability box for game 2*R* could lead

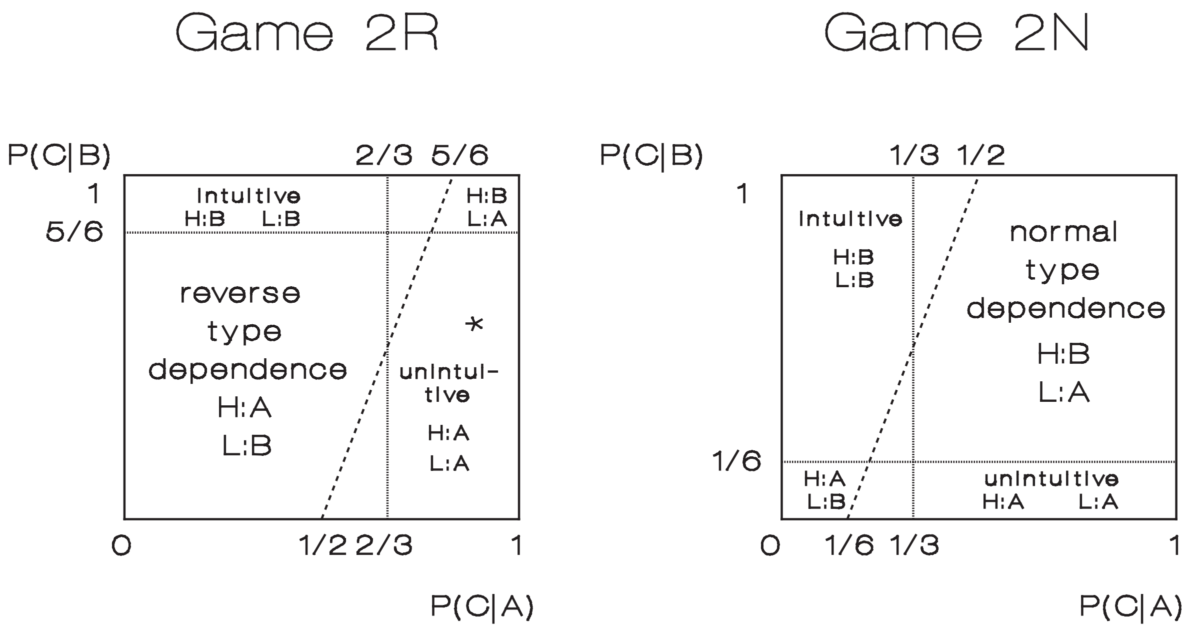

Figure 3. Optimal Signals by Type

Key: *T*:*S* (optimal signal is *S* for type-*T*)

to the unintuitive equilibrium in which the *A* signals are sent by both types, and prior beliefs in subsequent matchings may then lie in the southeast part of the left-hand box in figure 3. The boundaries of the regions in this box would be altered by non-neutral risk attitudes, but the basic relationships between the areas of the regions would persist.[18]

As a control, consider game 2*N* on the right side of figure 2. The control was obtained from game 2*R* by lowering the type-*H* payoffs for the *AC* and *AD* outcomes from 160 to 40, and by lowering the type-*L* payoffs for the *BC* and *BD* outcomes from 130 to 70. These reductions made the *A* signal much less attractive for the type-*H* proponent and made the *B* signal much less attractive for the type-*L* proponent. As a result, the horizontal dotted line that divides the decision regions in the right-hand box in figure 3 falls from the original height of 5/6 to a height of 1/6, and the vertical dotted line moves half-way towards the left side of the box.

[18] To analyze the effects of risk aversion on the positions of the lines that divide the probability box, consider the extensive form for game 2*R* on the left side of figure 2. Signal *A* involves no risk for a type-*H* proponent, so an increase in risk aversion will increase the (*H*:*A*) area in the probability box on the left side of figure 3. Thus risk aversion causes the horizontal dotted line to shift upward to levels above a height of 5/6. The analysis of risk aversion for type-*L* proponents is similar; since signal *A* is more risky than the safe signal *B*, an increase in risk aversion would expand the (*L*:*B*) region, and thereby shift the vertical dotted line to the right for game 2*R*. To summarize, these calculations imply that risk aversion would expand the reverse-type-dependence area.

Consequently, the area with reverse type dependence is reduced, and the northeast area with the normal type dependence is enlarged, which is the reason for the "*N*" designation with the game number.[19] As before, the control contains two pooling equilibria, and the one involving the A signals is ruled out by equilibrium dominance.

Treatment 3 was divided into three parts, as indicated in the relevant column of table 1. Game 2*N* was used in part a, while game 2*R* was used in parts b and c, with payoffs in pennies. Roles were reversed after each part. The labeling for the different decisions was altered after the first part.[20]

The results for treatment 3 are shown at the top of table 4. The summary data for the control treatment (part a) were roughly consistent with the intuitive equilibrium, as can be seen from the high incidence of (B,C) outcomes in the relevant part of table 4. For the 2*R* design used in parts b and c, reverse type dependence was strong, and indeed nearly perfect in part c, where all type-*H* proponents chose signal A, and 7 out of 8 type-*L* proponents chose signal B. Notably, the pattern of signals in part c was quite different from the pattern of pooled B signals predicted by the intuitive equilibrium and observed earlier in treatment 1. But neither was there a dominant tendency to choose A, the signal consistent with the other equilibrium.

Treatment 4

Treatment 4 was based on game 3*N*, used in part a, and game 3*R*, used in parts b and c (with role reversal). Games 3*N* and 3*R* are represented by the asterisk change shown in figure 2. In addition, in game 3*N* all respondents' payoffs were reduced by 25 from the levels shown in the figure, and all type-*H* proponents' payoffs were reduced by 10.[21] The type dependence

---

[19] It is straightforward to show that the analogous probability box figure for the game 1 parameters has no reverse-type dependence area at all, i.e. there is *no* combination of $P(C|A)$ and $P(C|B)$ for which the inequality in (1) is satisfied for type H but not for type L in game 1. Moreover, the area with normal type dependence covers more than two-thirds of the probability box for game 1. We found that it is possible to reduce the normal type-dependence area, but it is not possible to eliminate this area without changing the basic equilibrium configuration.

[20] The procedures of treatments 3, 4, and 5 differed from the others in that subjects were asked to record nonbinding "planned decisions" (type-contingent signals or signal-contingent responses) at the beginning of each matching. Subjects were told that the Planning Decision Sheets would remain at their desks and would not affect earnings. Decisions were often inconsistent with plans. In one of the sessions for treatment 5 described below, about 40% of the plans that could have been reversed were reversed. There was no pattern in the nature of reversals, nor was there a tendency for reversals to diminish over time. Therefore, we do not report planned decisions. A second procedural difference was that subjects were asked to "explain your objectives" after the experiment ended. Commonly mentioned objectives were maximization of earnings and minimization of risk.

[21] These changes, together with the changes in labeling of decisions on subjects' decision sheets, were made to reduce carry-over effects of behavior in game 3*N* on the subsequent play in game 3*R* in treatment 4.

Table 4. Data for Treatments 3-5 by Part

Treatment 3:

part a — game 2*N*

	C	*D*
A	2 (0,2)	2 (0,2)
B	11 (8,3)	1 (1,0)

part b — game 2*R*

	C	*D*
A	6 (4,2)	2 (2,0)
B	4 (1,3)	4 (0,4)

part c — game 2*R*

	C	*D*
A	8 (7,1)	1 (0,1)
B	4 (0,4)	3 (0,3)

Treatment 4:

part a — game 3*N*

	C	*D*
A	1 (0,1)	1 (0,1)
B	14 (11,3)	0

part b — game 3*R*

	C	*D*
A	6 (6,0)	3 (2,1)
B	7 (6,1)	0

part c — game 3*R*

	C	*D*
A	3 (2,1)	2 (2,0)
B	10 (9,1)	1 (1,0)

Treatment 5:

part a — game 3*R* (with 6 matchings)

	C	*D*
A	35 (26,9)	6 (5,1)
B	14 (12,2)	17 (8,9)

part b — game 3*R* (with 6 matchings)

	C	*D*
A	49 (36,13)	7 (5,2)
B	9 (3,6)	7 (4,3)

Key: N (N_H, N_L) indicates N outcomes, N_H with type-*H* proponents and N_L with type-*L* proponents.

areas are determined by the positively sloped dashed lines in the boxes in figure 3, but the configuration of intuitive and unintuitive equilibria is unchanged. The data for game 3*N* in part a, which are summarized in table 4, yielded a pattern that is very close to the pattern implied by the intuitive equilibrium. The results were mixed, and there was some tendency toward reverse type dependence. The data pattern in the final part of treatment 4 is mixed, but if anything it seems to be more consistent with the intuitive equilibrium.

Treatment 5

Although the results for treatment 4 were, at best, inconclusive, we were encouraged by the strong reverse type dependence and disequilibrium behavior observed in parts b and c of treatment 3. A possible interpretation of the fact that signal *A* did not predominate for late matchings of game 2*R*, as it would in the unintuitive equilibrium, was that the treatment had not allowed for enough experience with this game. This led to the design of treatment 5, which used 12 subjects in each of the two sessions that were conducted.[22] With 12 subjects instead of 8, the number of matchings could be increased from 4 to 6 in each part, as indicated by the right-hand column of table 1. This treatment consisted of only two parts, both based on game 3*R*, with role reversal between parts.[23]

The summary data for treatment 5 are reported at the bottom of table 4. There was a tendency toward reverse type dependence, which became stronger in the second part where a type-*L* proponent was more than twice as likely as a type-*H* proponent to send the *B* signal. In the second part, the *A* signal was sent 78% of the time (85% for type-*H* proponents and 62% for type-*L* proponents). Therefore, global behavior corresponded more closely to the unintuitive equilibrium, since *A* was the predominant signal for both types, and deviations to signal *B* were more likely to be made by type-*L* proponents.

Although respondents only "punished" the *B* signal about half of the time, both types of proponents had an incentive to choose signal *A* when faced with the empirical distribution of responses. In the second part, the empirical proportions of $(C|A)$ and $(C|B)$ responses were .88 and .56 respectively. These proportions can be used in place of the prior probabilities in inequality (1) to show that a type-*H* proponent would have earned an expected payoff of 160 by sending signal *A*, as opposed to 114 for signal *B*. Similarly, a type-*L* proponent would have

---

[22] We were also encouraged by a referee report that suggested the same design considerations that had motivated our choice of payoffs for game 2*R*.

[23] Therefore, there were 12 matchings as in earlier sessions, but there was not enough time to include a control (game 3*N*).

earned an expected profit of 168 by sending signal *A*, as opposed to 134 for signal *B*. The empirical response proportions for the second part of this treatment are represented by the asterisk at (.88, .56) in the right-hand box in figure 3.

Although these data do not reveal a perfect convergence to the unintuitive equilibrium, there is no support for the other equilibrium, i.e. the only one that is not ruled out by equilibrium dominance.

V. Conclusion

Many recent developments in game theory are motivated by a consideration of what is reasonable or intuitive in the context of a specific example, and the data from treatments presented here may serve as a useful supplement to the theorists' intuition. In each treatment, subjects were matched with others in a series of two-stage signaling games. Each game has two sequential equilibrium outcomes, one of which can be ruled out with the Cho and Kreps equilibrium dominance criterion, or with stronger refinements such as strategic stability. Games 1 and 1' were variations of the structure of an example used to motivate the Cho and Kreps (1987) discussion of equilibrium dominance. The observed outcomes generally corresponded to the "intuitive" equilibrium that is not ruled out by equilibrium dominance, and this correspondence was striking for matchings in which the players had some experience with different opponents. In addition, deviant signals were generally made by type-*L* proponents, which corresponds to the out-of-equilibrium beliefs that support the intuitive equilibrium. The unintuitive equilibrium could not even be reliably induced with a non-binding announcement that specified the strategies yielding this outcome.

The theoretical refinements of the Nash concept being discussed here are based on restrictions of beliefs off of the equilibrium path. After studying the adjustment of decisions to the intuitive equilibrium in the standard signaling games, we were able to design an alternative signaling game in which disequilibrium decisions in early matchings tended to be consistent with the (out-of-equilibrium) beliefs that support the *unintuitive* equilibrium. The data for this treatment are much closer to the unintuitive equilibrium with beliefs that, although inconsistent with equilibrium dominance, are roughly consistent with behavior encountered during the adjustment phase. There is little or no support for the intuitive equilibrium in this context.

To put these results into perspective, recall that game theory typically does not provide good predictions of decisions in nontrivial games when subjects have had no experience with other players in the same game. Therefore, it is natural to consider whether the performance of the theory will improve when either i) subjects obtain experience with a sequence of different players, or ii) equilibrium play is suggested to the subjects by a coordinator. The latter case is relevant for the design of incentive mechanisms; announcement effects are studied in Brandts and MacLeod (1989). With one exception, we focus on the former question, which is more relevant in economic applications where legal restrictions or communication problems preclude a coordinated effort to implement an equilibrium. Our experimental results indicate that equilibrium dominance and stronger refinements must be modified if they are to be useful in this context. The main implication of these results is that a positive theory of equilibrium selection in games should be based on properties of the adjustment process.

References

Banks, Jeffrey S., Camerer, Colin F. and Porter, David "Experimental Tests of Nash Refinements in Signaling Games," unpublished manuscript, May 1988, forthcoming in *Journal of Games and Economic Behavior*.

Banks, Jeffrey S. and Sobel, Joel, "Equilibrium Selection in Signaling Games," *Econometrica*, May 1987, 55, 647-61.

Brandts, Jordi and Holt, Charles A. "An Experimental Test of Equilibrium Dominance in Signaling Games," working paper, University of Virginia, December 1989.

Brandts, Jordi and MacLeod, Bentley, "Equilibrium Selection in Experimental Games with Recommended Play," unpublished draft, Universitat Autonoma de Barcelona, October 1989.

Camerer, Colin F. and Weigelt, Keith, "Experimental Tests of a Sequential Equilibrium Reputation Model," *Econometrica*, January 1988, 56, 1-36.

Cho, In-Koo and Kreps, David M., "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, May 1987, 102, 179-221.

Grossman, Sanford J., "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics*, December 1981, 24, 461-83.

Grossman, Sanford J. and Perry, Motty, "Perfect Sequential Equilibrium," *Journal of Economic Theory*, June 1986, 39, 97-119.

Harris, Milton and Raviv, Autur, "A Sequential Signaling Model of Convertible Debt Call Policy," *Journal of Finance*, December 1985, 40, 1263-81.

Holt, Charles A. and Scheffman, David, "Strategic Behavior and Antitrust," in James Meehan and Robert Larner, eds., *Economics and Antitrust Policy*, Westport Conn.: Greenwood Press, 1989, 39-82.

Kohlberg, Elon and Mertens, Jean-Francois, "On the Strategic Stability of Equilibria," *Econometrica*, September 1986, 54, 1003-37.

Kreps, David M. and Wilson, Robert, "Sequential Equilibrium," *Econometrica*, July 1982, 50, 863-94.

Milgrom, Paul and Roberts, John, "Price and Advertising Signals of Product Quality," *Journal of Political Economy*, August 1986, 94, 796-821.

Miller, Ross M. and Plott, Charles R., "Product Quality Signalling in Experimental Markets," *Econometrica*, July 1985, 53, 837-72.

Ochs, Jack and Roth, Alvin E., "An Experimental Study of Sequential Bargaining," *American Economic Review*, June 1989, 79, 355-84.

Riley, John, "Informational Equilibrium," *Econometrica*, March 1979, 47, 331-59.

Spence, Michael, "Job Market Signaling," *Quarterly Journal of Economics*, August 1973, 87, 355-74.

van Huyck, John B., Gillette, Ann B. and Battalio, Raymond C., "Credible Assignments in Non-Cooperative Games," working paper 88-33, Department of Economics, Texas A&M University, November 1988.