Journal of Economic Behavior & Organization 128 (2016) 195–208

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Forward induction: Thinking and behavior☆

Piotr Evdokimov[a], Aldo Rustichini[b,*]

[a] Department of Economics, ITAM, Mexico City, Mexico
[b] Department of Economics, University of Minnesota, Minneapolis, United States

ARTICLE INFO

Article history:
Received 4 August 2015
Received in revised form 1 May 2016
Accepted 2 May 2016
Available online 21 May 2016

Keywords:
Forward induction
Talk aloud protocols
First mover advantage

ABSTRACT

Forward induction (FI) thinking is a theoretical concept in the Nash refinement literature which suggests that earlier moves by a player may communicate his future intentions to other players in the game. Whether and how much players use FI in the laboratory is still an open question. We designed an experiment in which detailed reports were elicited from participants playing a battle of the sexes game with an outside option. Many of the reports show an excellent understanding of FI, and such reports are associated more strongly with FI-like behavior than reports consistent with first mover advantage and other reasoning processes. We find that a small fraction of subjects understands FI but lacks confidence in others. We also explore individual differences in behavior. Our results suggest that FI is relevant for explaining behavior in games.

1. Introduction

Forward induction (FI) is the notion that what an individual does in an early stage of a multi-stage interaction contains information about what he or she will do later. This information can be used strategically by others to facilitate coordination. FI was introduced to economics by Kohlberg and Mertens (1986) and found several successful applications, particularly in models of industrial organization (Ponssard, 1991; Ben-Porath and Dekel, 1992; Bagwell and Ramey, 1996). In the Bagwell and Ramey (1996) model, two firms, an entrant and an incumbent, make sequential investment capacity decisions in the presence of multiple equilibria, with the entrant moving first. The incumbent, assuming that the entrant is rational, deduces the entrant's post-entry production from the latter's capacity investment. If the capacity commitment is such that the entrant is not able to recoup its investment with a market sharing equilibrium, the incumbent concludes that the entrant will produce at natural monopoly levels and shuts down.

Experimental evidence supporting the hypothesis that players use FI thinking has been mixed. Bagwell and Ramey (1996) has been tested in the laboratory (Brandts et al., 2007). Other experiments on FI include Cooper et al. (1993), which studied the effect of introducing an outside option in a battle of the sexes game, and Huck and Müller (2005), which allowed subjects an opportunity to "burn money." All of these studies argued that while subjects often coordinate on equilibria selected by FI, other factors, such as asymmetries between the players, play a significant role in determining behavior. Summarizing the state of the literature, Samuelson (2005) reported that "the experimental evidence has not been particularly supportive

☆ We thank the participants of the University of Minnesota Mathematical Economics Workshop, the 2012 Fourth World Congress of the Game Theory Society, the 2013 ESA World Meetings, and the 2013 Midwest Theory Meetings for helpful comments and suggestions. Piotr Evdokimov acknowledges the financial support of Asociación Mexicana de Cultura. Aldo Rustichini thanks the NSF grant SES-1061817.

* Corresponding author. Tel.: +1 6126254816.
E-mail addresses: pevdokim@gmail.com (P. Evdokimov), aldo.rustichini@gmail.com, arust@econ.umn.edu (A. Rustichini).

http://dx.doi.org/10.1016/j.jebo.2016.05.002

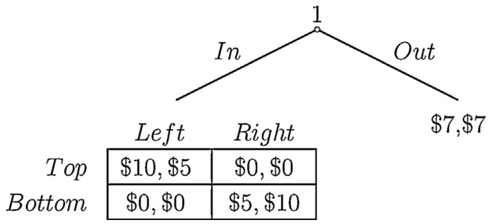

Fig. 1. The game used in the experiment.

of forward induction, suggesting that theories based on forward induction could well be reconsidered." This suggestion notwithstanding, developments in the theoretical literature on FI have anything but slowed.[1] It seems that the jury on FI is still out.

We contribute to the experimental literature on FI by incorporating a new source of data: subjects' elicited beliefs as expressed in reports of a relatively free-form nature. All previous experimental studies of FI have focused on participants' choices, leaving unsettled the question of whether FI thinking is actually used. Our experiment is the first to address the issue directly. We designed it around a battle of the sexes game with an outside option illustrated in Fig. 1. This is the simplest game in which FI makes a prediction.[2] Participants were instructed to play it in anonymous, randomly determined matches. In several rounds, they answered a series of questions aimed at assessing one's understanding of the FI argument. Their answers to these questions were scored by independent evaluators following an incentivized procedure described in Xiao and Houser (2005).[3]

We organize our results around two major hypotheses. The first of these deals with subjects' ability to use FI thinking:

Hypothesis 1. Subjects use FI thinking in the game illustrated in Fig. 1.

As the most likely alternative to Hypothesis 1, we take the focal point-based reasoning discussed in Cooper et al. (1993) and Huck and Müller (2005).[4] While previous experiments cannot rule out the possibility that subjects use such focal point-based reasoning in treatments where FI has no bite (the same game with a $1 outside option), ours addresses this issue directly:

Hypothesis 2. FI thinking is associated with more FI behavior than using the asymmetry in the game to coordinate on a focal point.

We find support for both hypotheses in the data. First, a substantial fraction of reports—30%—show an excellent understanding of FI. 33% of reports are consistent with FI as well as focal point-based reasoning, and the remaining reports show no evidence of FI. About 50% of participants produce at least one report showing a strong understanding of FI reasoning, which provides at least partial evidence for Hypothesis 1. Second, reports showing strong evidence of FI are associated with more FI-like choices in the game. This is true in rounds when reports were provided, as well as those when they were not. In the former case, a report showing strong evidence of FI increased the probability of FI-like behavior by about 36%, while a report consistent with FI as well as focal point reasoning increased this probability by about 19%. These two effect sizes are significantly different. Thus, our results show a relationship between FI thinking and behavior, as suggested by Hypothesis 2.

An unconfident subject might report that the second mover will choose *Right* if the subgame is played, but that they themselves would use FI and choose *Left* in the role of the second mover. We find that unconfident subjects are indeed present in the data. Our results support the hypothesis that confidence in FI can be encouraged in two ways. First, subjects are more likely to produce a report consistent with FI (and exhibit FI behavior) if they observe more of their partners choose *In*. Second, FI behavior is significantly more likely in periods with belief elicitation, suggesting that more careful thinking about the game affects participants' thinking and behavior.

Although this raises the possibility that understanding of FI was encouraged by our belief elicitation procedure, the results suggest that this is unlikely. If a participant understands FI, he should not lose this understanding in a later period. We find, however, that the behavioral effect of belief elicitation is transient. Thus, FI behavior is less likely in periods following belief

[1] See, for example, Govindan and Wilson (2009), Battigalli and Friedenberg (2012), Man (2012), and Müller (2014).

[2] The first mover chooses between an outside option of $7 and playing a battle of the sexes game, in which the first mover's favored payoff is $10. Under these parameters, the first mover's decision to play in the subgame contains the following implicit message, quoted from Kohlberg and Mertens (1986): "Look, I had the opportunity to get [$7] for sure, and nevertheless I decided to play in this subgame, and my move is already made. And we both know that you can no longer talk to me, because we are in the game, and my move is made. So think now well, and make your decision." For the second mover to "think well" is to realize that any strategy other than a sure choice of *Top* is inconsistent with a rational first mover's decision to play in the subgame. The best response of a second mover who "thinks well," therefore, is to choose *Left* whenever he is given the opportunity to play. Accounting for the above, a rational first mover chooses *In* in the first stage of the game, and *Top* in the second.

[3] Specifically, we hired outside evaluators, instructed them how to classify the free-form reports, and incentivized performance using a coordination game in which a rater is paid if his classification agreed with that of the majority. The details of the rating procedure are described in Section 3.

[4] It is argued in Cooper et al. (1993) that offering an outside option to the row player makes his preferred equilibrium focal.

elicitation. Second, belief elicitation only significantly affected the behavior of subjects who showed an understanding of FI in one of their reports. This is consistent with the hypothesis that these subjects lacked confidence in FI but were encouraged in periods when they and their partners thought more carefully about the game.

We conclude that FI thinking is commonly used in a battle of the sexes game with an outside option, and that such thinking significantly predicts FI behavior. Unlike other experimental studies, ours provides *direct* evidence of FI thinking. We also find that some subjects lack confidence in FI, and that these subjects can be encouraged by observation and deliberation. Our results, therefore, suggest that FI is a useful theory for explaining behavior in games.

2. Experimental design

We now outline the experimental design. Section 2.1 provides the details of the protocol; Section 2.2 focuses on our choice of feedback structure; Section 2.3 describes the questions participants answered, and Section 2.4 motivates and discusses our treatments. A description of how the reports were categorized is deferred to Section 3.

2.1. Protocol

The experiment included four treatments (labeled T1–T4), all run in the Social and Behavioral Sciences Laboratory at University of Minnesota in the fall semesters of 2011 (T1 and T2) and 2012 (T3 and T4). The treatments were programmed with z-Tree (Fischbacher, 2007) and used University of Minnesota undergraduate students as participants. Upon signing a consent form, each subject was seated at a computer terminal and given a paper copy of the instructions, the complete text of which is included in the online appendix. Subjects first read the instructions, then completed a multiple choice quiz on the details of the game, displayed on a computer screen.[5] Completing the quiz correctly on the first try resulted in a small ($1) reward. Then, forty (in T1 and T2) or eight (in T3 and T4) rounds of the game were played.

Half of the participants in each session played as first movers in even and second movers in odd rounds; for the other half, the reverse was true, and this information was known to all subjects. Hence, every session was split into two groups, and, throughout the experiment, participants that started out as first movers were matched with those that started out in the other role. Subjects were not told how many times they will play the game, or *when* they will have to report on their thinking, but they were informed that they will "play the game a number of times," that their partner will be randomly selected in every round, and that they will answer questions about their thinking at some points in the experiment. Some of these answers were incentivized, as described in Section 2.3. Second movers were always informed of their opponent's first stage choice, that is, whether their partner had chosen *In* or *Out*, but no other feedback was given. Participants in T1 and T2 were paid on the basis of the outcome of four rounds, and those in T3 and T4 were paid for the outcome of two. These rounds were drawn at random at the end of the session in addition to a $5 show up fee. Participants received their earnings from the games at the end of the session, and, four to five weeks later, a check in the mail for their reports.[6]

2.2. Feedback structure

As described above, our participants were provided with limited feedback about the game: Second movers only knew if the subgame was played, and first movers had no feedback at all. We consider this a crucial element of the design. Kahneman (2011, p. 203), for example, notes that people "are prone to assess the quality of a decision not by whether the process was sound but by whether its outcome was good and bad." Because we viewed FI as a kind of sound process described by Kahneman, we needed to ensure that it is not confounded by observing the decisions of the other players. We expected this feature of our design to have consequences on behavior. On the one hand, it is argued by Rick and Weber (2010) that "withholding feedback encourages deeper thinking about the game." On the other, not knowing what their opponents choose in the subgame, subjects who don't understand FI, or lack confidence in the ability of others to use FI reasoning, are likely to assume the worst (namely, a choice of *Right*) and opt for the outside option. Thus, we expect more outside options being chosen in an experiment with little opportunity for learning.

2.3. Questions

We now describe the questions answered by first movers in rounds when reports were provided. For reasons described in Section 3, the second movers' answers were not scored by the evaluators. They were, however, scored by the researchers and played a role in determining subjects' payments, as discussed in more detail below. We include the second movers' questions in the online appendix below.

Before the decision to go *In* or *Out* was made, each first mover answered F1–F3:

F1. If you go in, what move will the second player make?

[5] The quiz questions are included in the online appendix.
[6] The time interval was necessary to rate the reports.

Possible answers: *Left* for sure, 90/10 *Left/Right*, 80/20 *Left/Right*,. . ., *Right* for sure.

F2. Why will the second mover make this move?

F3. If you go in, what will the second mover think you will do?

Possible answers: *Top* for sure, 90/10 *Top/Bottom*, 80/20 *Top/Bottom*,. . ., *Bottom* for sure.

If *Out* was chosen in the first stage, the first mover saw a screen that said: "We are now informing the second mover that you chose Out. Press OK to continue." If *In* was chosen, the first mover had to choose their next move and answer F4.

F4. Why did you make this move?

We provided incentives for multiple choice and typed answers with monetary rewards. Where participants guessed actions that were eventually realized, we used the quadratic scoring rule to calculate their payment (denoted by π below).[7] For example, if a first mover chose *In*, they were paid in cents for their answer to F1 according to

$$\pi_{S1} = \begin{cases} 100 - (Prob_{F1}(Right))^2 & \text{if the second mover chose } Left \\ 100 - (Prob_{F1}(Left))^2 & \text{if the second mover chose } Right \end{cases},$$

where $Prob_{F1}(Right)$ is the probability assigned to the second mover choosing *Right* by the first mover's answer to F1. This scoring rule ensures that a risk-neutral decision maker has the incentive to report truthfully. To facilitate subjects' understanding of the scoring rule, the instructions included several examples of guesses, choices being guessed, and payments for the guesses, as well as a statement encouraging subjects to report truthfully.[8]

If a player made a guess about what their opponent is thinking, they were fined with a square of the distance between their and their opponent's reports. For example, in question S3, second movers were asked to guess what their partner will do in the case the subgame is played (see Section A). In question F3, first movers were asked to guess what their partner will think they (the first movers) will do, also in the case the subgame is played. If the first mover chose *In* (i.e., if the subgame was played), the first mover's payment for F3 was calculated as

$$\pi_{F3} = 100 - (S3_{Top} - F3_{Top})^2,$$

where $S3_{Top}$ is the probability assigned by the second mover to their partner choosing *Top* in their answer to S3. This scoring rule ensures that a subject with point beliefs has the incentive to report his conjecture, while a subject maximizing expected earnings with a non-trivial belief set has the incentive to report his expectation of the partner's report.[9] The instructions explained this scoring rule to the subjects and included a statement that it is in the subjects' best interests to report truthfully.

To provide appropriate incentives for the written reports, we told participants that each statement describing *a partner's* thoughts or behavior would be evaluated by us, and classified into a category. The same procedure would be applied to the partner's description of *his own* reasoning process, and only the former subject (the one describing his partner's thoughts) would be paid in the event the categories matched. A statement describing one's own behavior was not incentivized. The specific instructions were:

> For each question in which you explain what the other player is doing (or thinking), this player will have a corresponding question in which they have to explain their own behavior (or thoughts). We will evaluate your answer by placing it into one of several categories, and do the same for the answer the other player provides. If the categories match, we will pay you $1. For example, if your explanation of why the other player will behave in a particular way matches that player's explanation of their behavior, you get $1. If your explanations do not match, we will pay you $0. If one of you does not provide a short answer, we will pay you $0.

Moreover, the subjects were instructed as follows:

> We will not pay you additional money for answers to questions about your own thoughts and behavior, but obtaining your considered opinion is important for our study, so please be as detailed as possible in your answers.

Note that when one player is making a guess about another's thoughts, only the former individual gets paid in the case of a match. The other player's report, in this case, is either not incentivized (if own behavior is explained), or incentivized

[7] Although there is some evidence that these rules are associated with misreporting (Armantier and Treich, 2013), there is also evidence that the distortions in reports are small or non-existent (Sonnemans and Offerman, 2001).

[8] Specifically, the statement was: "The closer your guess is to their choice, the more money you make." See the online appendix for the instructions.

[9] If $f(x)$ is a subject's subjective probability distribution over *stated* beliefs of his partner, the minimization problem is:

$$\min_{p \in [0,\ 1]} \int (x-p)^2 f(x)dx.$$

It is trivial to check that $p = \int x f(x)dx$ satisfies the first and second order conditions.

by paying this player more if his guess of the partner's behavior is closer to the truth. Thus, our design does not induce a coordination game with multiple equilibria in statements, where a player has an incentive to state a less sophisticated explanation than his true belief if he thinks his matched player might think in a simpler way, just to coordinate with what he thinks the other player is thinking. The player would only want to do this if he is altruistic.

2.4. Treatments

In treatment **T1**, subjects answered the questions described in Section 2.3 in rounds 5, 6, 11, 12, 17, and 18 of the game. In treatment **T2**, the questions were answered in rounds 25, 26, 31, 32, 37, and 38. Thus, each subject in these treatments had three opportunities each to describe their thinking as first and second mover. In addition to assessing how prevalent FI thinking is and how it relates to behavior, these two treatments allowed us to see whether FI thinking becomes more likely with experience. A cross-treatment comparison allowed us to see whether participants were more likely to demonstrate FI thinking after twenty rounds of playing the game.

One may fail to understand the logic of **FI** altogether, or doubt that one's opponent thinks about the game in such terms. The two treatments described below were designed to determine whether confidence in one's partner played a role in shaping participants' reported beliefs. **T3** and **T4** each consisted of eight rounds of play, with questions answered in the fifth and sixth rounds. Both treatments differed from T1 in two respects: session length and payment structure. Thus, subjects were paid for the outcomes of *two* games in T3 and T4, whereas in T1 they were paid for four. Treatment T4 had the additional feature that first movers choosing *Out* were asked the following three questions:

F5. What would the second mover choose if instead you had moved in?
 Possible answers: *Left* for sure, 90/10 *Left/Right*, 80/20 *Left/Right*,. . ., *Right* for sure.
F6. Why would the second mover have made this choice?
F7. Put yourself for a moment in the role of your partner. Suppose you were the second mover, and you saw the first mover move in. What would YOU think in that case?

Some subjects may understand FI but lack confidence in others. Under this hypothesis, more evidence of FI thinking should be exhibited in T4 than in T3 if and only if the hypothetical reports are taken into consideration. By looking at subjects who showed an understanding of FI only in their hypothetical reports, T4 also allows us estimate the percentage of unconfident subjects in the data.

3. Rating procedure

The reports were rated by outside evaluators.[10] Ten evaluators were recruited in January 2013 with an e-mail to PhD students at the School of Physics at the University of Minnesota. Students of physics were chosen for two reasons. First, they were unacquainted with the researchers, so their reports were unlikely to be biased. Second, the quantitative nature of their discipline increased the likelihood that the raters will understand the notion of FI sufficiently well to provide reasonable ratings. For their participation, the raters were paid $15 per hour. In addition, four reports were selected at random at the end of the session, and each rater was paid an additional $10 for each instance of their classification agreeing with that of the majority.[11]

Due to the subjective nature of the procedure, it was important that the same raters classified every report in the data set. Moreover, the reports needed to be rated in a single session. Allowing the raters to go home and return at a different date to finish the scoring procedure carries a cost of losing control over what the raters say to each other. Thus, for instance, the raters might have agreed on a scheme that maximizes payment at the expense of classification accuracy. On the other hand, 389 reports were provided by first movers alone in T1–T4.[12] We estimated the rating procedure to take at least 3 h even if reports of second movers are not shown to the raters. Thus, for the sake of time and simplicity, only reports of first movers were scored by the outside evaluators.

To pay the subjects for their reports, we classified all reports of second movers in treatments T3 and T4 as potentially showing evidence of FI ("MAYBE") reports. This is an intermediate category that were taken to be most neutral. Note that the raters' payments did not depend on these uniform classifications: The coordination game played by the raters only involved first movers' reports, and the raters were fully aware of this. The subjects in T3 and T4 were paid with a check in the mail

[10] In addition, a portion of the reports was scored by the authors of the paper, as in Charness and Dufwenberg (2006). Our ratings agree with those of the physics students to a large extent; Section 5 discusses this in more detail.

[11] One rater asked if such a scheme incentivizes reporting what one believes his peers to believe, rather than one's personal beliefs. Our answer was that our true interest lied in each rater's *own* beliefs about the reports.

[12] In T1 and T2, 108 subjects had three opportunities to report on their thinking. In T3 and T4, 122 subjects had one opportunity each. The maximum possible number of reports in the experiment is $108 \times 3 + 122 = 446$. We have fewer reports because some subjects failed to answer the questions.

based on our (uniform) ratings as well as the ratings of the raters.[13] In T1 and T2, all reports were scored by us a year before the evaluators were hired, and it was on the basis of these ratings that the subjects were paid.

After entering the laboratory and signing a consent form, the raters were instructed by the experimenter using a PowerPoint presentation.[14] The presentation described the game, the structure of the experiment producing the report data (feedback, order of choices, questions answered by participants, etc.), and explained the notion of FI. It then gave examples of real reports provided by first movers that could be grouped in each of the three suggested categories: those showing evidence of understanding FI ("YES" reports), those *potentially* showing such evidence ("MAYBE" reports), and those *not* showing it ("NO" reports).

A report consisted of the first mover's answers to F1, F2, and F4 (the latter if available), as well as the player's choice in the game. The raters were told that a report cannot be judged to show understanding of FI without an explanation of behavior (i.e., an answer to F2 or F4) present, since a belief that the second mover will choose *Left* in the subgame with 100% certainty and, in turn, believe that the first mover will chose *Top* can be supported by thought processes other than FI. Thus, the raters were told to pay attention to the text and only look at the answer to F1 or the subject's choice if something in a typed response needs to be clarified. Importantly, the raters were also told that a player choosing *Out* can understand FI, and that their job as raters is to identify such understanding—not its reflection in behavior.

A YES report was explained to be one that explicitly stated that giving up an outside option of $7 implies that the first mover is looking to make $10, choose *Top*, or both, or strongly hinted at the understanding of this fact. The example given to the raters contained the following sentence (here, as in other examples, original spelling and grammar are preserved):

> I chose to go in so I have sent the message to the second mover that I intend to make more than the original $7, I can only do this if I choose top.

If the described reasoning process was consistent with understanding FI, but could have been used in a similar game with the outside option replaced by ($3, $3), the raters were told that the report belongs in the MAYBE category. Two examples of MAYBE reports were given, one of them replicated below.

F1. If you go in, what move will the second player make?
Answer: 80/20 *Left/Right*.
F2. Why will the second mover make this move?
Answer: I figure that the second mover will make this move because they will figure that I will choose top, because that way I make more money, and instead of not making any money, the second mover will most likely choose left and get $5.
F4. (regarding the choice of In, Top)
Answer: I chose top because that way I will make more money, but I am hoping that the second player will choose left because they know that I will choose top only because I can make more money and I hope they think that I am not being generous by choosing bottom for them to make $10 instead of me.

Finally, the raters were told that a NO report should be one clearly not consistent with understanding of FI. Two examples of such reports were given, one of them stating the probability of the second mover choosing *Right* in the subgame to be 70%, and containing the following passage.

> Realistically speaking, I think it's 50/50 in terms of what the smartest move is. Once you figure in the human emotion known as hope, preferences for certain options show up. So the idea behind my 70/30 is that hope can account for 20 percent change. I'm really just going on hunches. As for right, most people are right handed. So I put two and two together.

Following the instructions, the raters were told that they will be presented with the reports. They were also told that after all the reports are rated, the experiment will have a second, shorter, portion, and that the instructions to this half of the experiment will be given after everyone is done with the initial set of ratings. The reports were then presented through an E-Prime interface. A rater had a maximum of 40 seconds to rate each report. An untimed break was provided every 20 trials. During the break, the rater could rest, go to the bathroom, or, if desired, continue the scoring procedure.

After every rater finished with the first set of reports, a second set of instructions, describing T4, was presented. The raters were instructed that they will score 38 reports consisting of the answers to F5, F6 and F7, and that they should follow the criteria established previously. After the second set of reports was rated, four reports were chosen, and each rater's classifications were compared to that of the majority. The raters were then paid for their time and classification choices.

[13] As discussed in Section 2.3, the subjects were not made aware of all of the details of our rating procedure, knowing only that we would classify their reports in categories using an unspecified method. This omission is a necessary feature of our design, since a complete description of the classification procedure would also require an statement about forward induction reasoning.

[14] The slides are available in the online appendix.

Table 1
Summary of behavioral data. For T1 and T2, the table shows the overall averages across rounds (top row in each panel), as well as the likelihood in the first eight rounds (bottom row in each panel). Thus, second movers in T1 choose *Left* 86% of the time overall, and 77% of the time in the first eight rounds of the experiment. For T3 and T4, where subjects played the game for eight rounds, the table shows the overall averages in both rows of every panel.

	T1	T2	T3	T4	T1–T4 (all)
N	48	60	52	70	230
In as first mover (all rounds)	0.45	0.36	0.33	0.35	0.39
In as first mover (rounds 1–8)	0.39	0.33	0.33	0.35	0.34
Top as first mover in subgame (all rounds)	0.98	0.95	0.85	0.88	0.95
Top as first mover in subgame (rounds 1–8)	0.95	0.88	0.85	0.88	0.89
Left as second mover in subgame (all rounds)	0.86	0.84	0.74	0.65	0.82
Left as second mover in subgame (rounds 1–8)	0.77	0.85	0.74	0.65	0.74
FI behavior as first *and* second mover (all rounds)	0.43	0.34	0.25	0.26	0.36
FI behavior as first *and* second mover (rounds 1–8)	0.34	0.28	0.25	0.26	0.28

4. Results

We now turn to our experimental findings. In Section 4.1, we summarize participants' behavior and analyze their choices in the game. In Section 4.2, we estimate how many participants understood FI and study how likely this kind of thinking is to be learned. In Section 4.3, we quantify the number of subjects who understood FI but lacked confidence in the understanding of others. Section 4.4 studies the relationship between reported thinking and behavior.

4.1. Behavior

A total of 230 subjects participated in the experiment, with three sessions each in T1 and T2 and four sessions each in T3 and T4. The session sizes were 10, 18, and 20 subjects in T1; 18, 20, and 22 subjects in T2; 12, 12, 14, and 14 subjects in T3; and 14, 16, 18, and 22 subjects in T4. Table 1 shows the breakdown of the overall sample size by treatment, as well as a summary of choice data. The fraction of first movers choosing *In* fluctuates from 33% in T3 to 45% in T1, and the overall average (39%) is substantially smaller than the 80% observed by Cooper et al. (1993). We take this to be a consequence of the structure of feedback in our design. Thus, subjects who do not observe their partner's choice in the subgame (and hence lack evidence of second movers choosing *Left*) are more likely to choose the outside option than those who do.[15]

Table 1 also tabulates behaviors of first and second movers in the case the subgame was played. Overall, the probability of first movers choosing *Top* is 95%, while that of second movers choosing *Left* is 82%. Second movers in our experiment chose the FI strategy less often than those in Cooper et al. (1993), presumably because exclusion of feedback boosts their optimism about attaining the more desired outcome. Comparing behavior in the first eight rounds across treatments, we find no differences in likelihoods of first movers choosing *In* ($P = 0.553$; Fisher's exact test) or *Top* ($P = 0.291$; Fisher's exact test).[16] Although we find a treatment effect on the choices of second movers ($P < 0.05$; Fisher's exact test), this effect loses significance when T2 is excluded from the analysis ($P = 0.202$; Fisher's exact test). We conclude that subjects in different treatments behaved similarly in early rounds of the game.

We take the view that thinking about the game is a unitary process. If a player understands FI as a first mover, he should understand it as a second mover, as well. While the latter sort of understanding requires a smaller degree of strategic sophistication, our own preliminary investigations suggest that this is not the case: We find as many subjects showing evidence of FI as first but not second movers as those demonstrating the opposite pattern.[17] Thus, in the remainder of our behavioral analysis, we focus on "periods" of play. A period is a pair of two consecutive rounds (e.g., period 1 consists of rounds 1 and 2). Using this definition, participants in T1 answered questions about their thinking in periods 3, 6, and 9; participants in T2 answered them in periods 13, 16, and 19, and participants in T3 and T4 answered them in period 3. Notice that every participant played once as a first mover and once as a second mover in every period.

We define a subject to be a *FI player* in a given period if, in this period, he chose *In* and *Top* as a first mover, and, provided that he was given an opportunity to play in the subgame, *Left* as a second mover. *FI player* should be read as shorthand for "a player exhibiting behavior consistent with FI as both the first and the second mover." We emphasize that being such a player does not imply an understanding of the FI argument. In Section 4.4, we study the relationship between thinking and behavior directly.

---

[15] Another difference between our and Cooper et al.'s design concerns structure of payment. In Cooper et al. (1993), participants played for points, and instead of our \$7, \$10 and \$5, they could earn 300, 600 and 200 points, respectively, in each game played; the probability of receiving a payment was calculated as the number of points earned divided by 1000. We chose dollar amounts for two reasons. First, Selten et al. (1999) showed that lotteries do a poor job at inducing risk neutrality. Second, we wanted to keep our instructions as simple as possible.

[16] Recall that the eighth round was final in T3 and T4.

[17] Specifically, we find that 61 subjects in T1 and T2 failed to provide a YES report. Of the remaining subjects, 22 provided YES reports in both roles; 12 provided it as second but not first movers, and 13 provided it only in the role of first movers. This suggests that if subjects are given multiple opportunities to report on their thinking, they are equally able to show evidence of FI in either role.

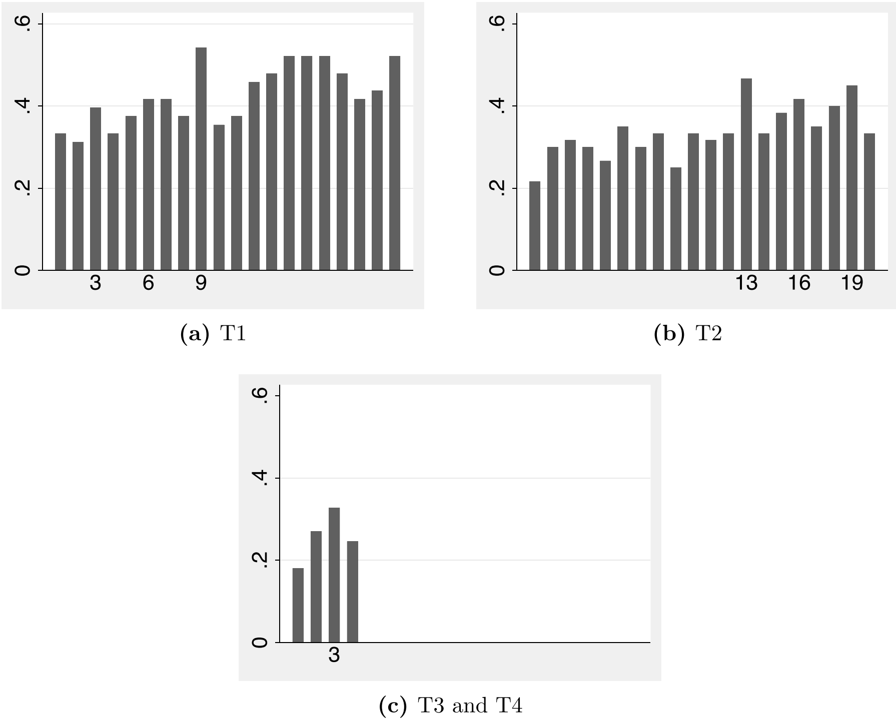

Fig. 2. The probabilities of displaying FI-like behavior in different treatments. Note the "spikes" in periods when participants answered questions about their thinking.

The bottom rows of Table 1 summarize the prevalence of FI-like behavior in T1–T4. Overall, behavior was consistent with being a FI player 36% time, and no significant treatment effect is found in the first four periods of the experiment ($P = 0.159$; Fisher's exact test). In Fig. 2, we plot the probabilities of being a FI player in each period of the game for participants in T1 (a), T2 (b) and T3 and T4 (c). The figure suggests that this probability went up in periods participants answered questions about their thinking. To confirm this observation, we estimate the following logit model:

$$P(FIPLAYER_{it}=1)=\frac{1}{1+\exp^{-1}(\beta_0+\beta_1 vr_{it}+\beta_2 t+\sum_{k=2}^{5}\beta_{3,k}Treatment_{ik})}, \quad (1)$$

where $vr_{it} = 1$ if i provided a verbal report in period t and 0 otherwise, and

$$FIPLAYER_{it}=\begin{cases}1 & \text{if, in period } t,\ i \text{ chose } (In, Top) \text{ as first and, given the chance, } Left \text{ as second mover}\\ 0 & \text{otherwise}\end{cases}.$$

We estimate this model with bootstrapped standard errors.[18] The results are reported in the left column of Table 2. We find that the marginal effect of period number is significant, suggesting that FI-like behavior is learned over time ($P < 0.001$). There are two reasons why this might be the case. First, additional experience might encourage deeper thinking about the game through introspection. Second, experience might have provided subjects with observations suggesting FI (or FI-like) thinking on the part of other players. We explore these possibilities by including a variable called OBS_{it} in the regression. This variable captures the number of times i observed his or her partner choose *In* in periods prior to t. The results are reported in the third column of Table 2. When both variables are included in the regression, the marginal effect of OBS_{it} is positive and significant ($P < 0.001$) while that on period number reverses its sign.[19] We interpret this to mean that most of the learning in the experiment is driven by observation rather than introspection.

[18] Though the number of sessions in our experiment is small, allowing for within-session correlations in standard errors has little effect on our conclusions. One interpretation is that the limited nature of feedback in our experiment made it difficult for subjects to learn from each other. We include the analysis using session-clustered standard errors in the online appendix.

[19] That the negative coefficient on period number is marginally significant in the third column of the table is likely a result of multicollinearity. Indeed, the correlation between period number and the OBS_{it} variable is close to 0.8. In addition a robustness check with session-clustered errors, the online appendix includes estimation results from a model with subject fixed effects. The period variable is not significant in either of the robustness checks when OBS_{it} is included as a covariate.

Table 2
The effect of learning and belief elicitation on behavior. FI behavior became more likely over time and in periods when verbal reports are provided.

	(1) FI behavior	(2) FI behavior	(3) FI behavior
Period no. (t)	0.007**** (0.002)		−0.005* (0.003)
No. of times partner chose In (OBS_{it})		0.023**** (0.003)	0.030**** (0.005)
Verbal report in period (vr_{it})	0.076**** (0.020)	0.078**** (0.020)	0.081**** (0.020)
Treatment fixed effects	Yes	Yes	Yes
Constant	Yes	Yes	Yes
Observations	2648	2648	2648

Marginal effects of logit regressions. Bootstrapped standard errors in parentheses.
* $p<0.10$.
**** $p<0.001$.

Table 3
Percentages of reports in different categories, as scored by outside evaluators.

	T1	T3	T4	Period 3 (all treatments)
YES	0.28	0.15	0.28	0.24
MAYBE	0.38	0.31	0.27	0.31
NO	0.35	0.54	0.45	0.45

(a) Reported thinking in period 3

	Period 3	Period 6	Period 9	T1 (all periods)
YES	0.28	0.37	0.37	0.34
MAYBE	0.38	0.32	0.34	0.34
NO	0.35	0.32	0.29	0.32

(b) Reported thinking in T1, across periods

	Period 13	Period 16	Period 19	T2 (all periods)
YES	0.34	0.33	0.33	0.34
MAYBE	0.38	0.31	0.37	0.36
NO	0.28	0.35	0.30	0.31

(c) Reported thinking in T2, across periods

In all columns of Table 2, the marginal effect of the elicitation dummy is positive and highly significant ($P<0.001$). One possibility is that the belief elicitation procedure improved players' understanding of FI. Another is that it increased players' confidence in other players. The data provides more support for the confidence channel for two reasons. First, the effect of belief elicitation on behavior was transient. i.e., FI behavior became *less likely* following the periods where participants answered questions.[20] If a participant understood FI, it is unlikely that he lost this understanding in a later period. Second, as we show in Section 4.4, the belief elicitation procedure had a significantly stronger effect on behavior of participants who showed an understanding of FI.

4.2. Thinking

Overall, the evaluators judged 30% of reports to show an understanding of FI, placing 33% of reports in the MAYBE and 37% in the NO category. We highlight this as our first major result:

Result 1. One third of the reports shows strong evidence of FI thinking.

Table 3 shows how these numbers break down across treatments. Recall that all treatments but T2 had subjects report on their thinking in the third period (T1, in addition to this, asked for reports in periods 6 and 9). It is apparent from the table that the fraction of period 3 reports judged to show evidence of FI fluctuates across treatments (from 15% in T3 to 28% in T1 and T4; see Table 3). Nevertheless, there is no significant treatment effect on how reports in this period were categorized ($P=0.255$; Fisher's exact test). In T1 and T2, where participants provided reports in multiple periods of the game, 34% of reports show an understanding of FI. The fraction of *participants* showing understanding of FI is even higher. Seven subjects provided no reports at all in these two treatments, and, of the remaining 101, 47% had at least one YES report.

[20] In Fig. 2, this can be seen as "spikes" in FI-like behavior. In the above-described regression, this can be inferred from the fact that learning is controlled for.

Table 4
The effect of feedback on the thinking of participants. Observing your partners choose *In* made a YES report more and a NO report less likely. Observations where no reports were provided are excluded from the sample.

	(1) YES report	(2) MAYBE report	(3) NO report
Period no. (t)	−0.008	−0.005	0.014
	(0.012)	(0.012)	(0.011)
No. of times partner chose In (OBS_{it})	0.035****	0.010	−0.046****
	(0.011)	(0.013)	(0.013)
Treatment fixed effects	Yes	Yes	Yes
Constant	Yes	Yes	Yes
Observations	389	389	389

Marginal effects of a multinomial logit regression.
Bootstrapped standard errors in parentheses.
**** $p < 0.001$.

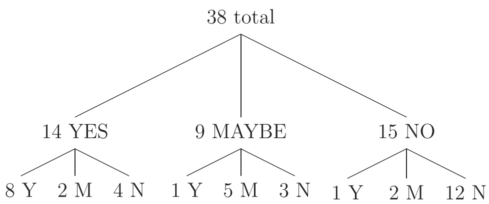

Fig. 3. The 38 auxiliary reports of T4 and the basic reports associated with them.

It's possible that more subjects understood FI in T1 and T2 because these treatments provided them with an opportunity to learn. The second and third panels of Table 3 show how reports were classified in different periods. Fisher's exact test finds no significant effect of period number on report categories in T1 ($P = 0.897$) or T2 ($P = 0.933$). Recall that reports were elicited early on (periods 3, 6, and 9) in T1 and later in T2 (period 13, 16 and 19). If experience with the game made FI thinking more likely, we should observe a significant difference between how reports are categorized in T1 and T2. However, no such difference is manifest in our data ($P = 0.988$; Fisher's exact test).[21]

One possibility is that FI is learned on the basis of observing other players, and not playing the game per se. To explore this hypothesis, we define p_{itj} as the probability of subject i producing a report of type j in period t and estimate the following multinomial logit model:

$$p_{itj} = \frac{\exp(\beta_{0j} + \beta_{1j}t + \beta_{2j}OBS_{it} + \sum_{k=2}^{5}\beta_{3jk}Treatment_{ik})}{1 + \sum_{j \in \{YES, MAYBE\}} \exp(\beta_{0j} + \beta_{1j}t + \beta_{2j}OBS_{it} + \sum_{k=2}^{5}\beta_{3jk}Treatment_{ik})}.$$

We estimated this model only for periods where participants provided reports, excluding observations where no report was provided from the sample.[22] The results are reported in Table 4. We find that observing one's partner choose *In* made a YES report more and a NO report less likely ($P < 0.001$ in both cases). This confirms the hypothesis that FI thinking can be encouraged by observation. The following section directly investigates the hypothesis that some subjects understood FI but lacked confidence to act upon this understanding.

4.3. Confidence in others

Recall that while we find that ≈50% of participants exhibited an understanding of FI in T1 and T2, the fraction of participants showing such understanding in T3 and T4 was smaller. One possibility is that a large fraction of participants (≈50%) understood FI in every treatment, but that these participants wavered in their confidence of other subjects' understanding. T4 was designed to address this hypothesis. In it, first movers choosing *Out* were asked to explain how *they* would have thought about the game as a second mover. Of the 38 reports, 14 were judged to show evidence of FI reasoning (Fig. 3). Of these 14, six were provided by participants judged to provide a MAYBE or NO report while they were playing the game. Thus, six out of 70 participants in T4, approximately 9%, understood FI but lacked the confidence to express this understanding while they were making their first stage decisions.

These participants, while few in number, explain the discrepancy between the percentages of players who understood FI in the longer (T1 and T2) and shorter (T3 and T4) treatments. If we only consider reports obtained in the course of the game,

[21] In light of the fact that FI-like behavior became more prevalent over time (Table 2), these findings may appear puzzling. We discuss the issue of learning further in Section 4.4.
[22] Including these observations has no qualitative effect on the results.

Table 5
Thinking and behavior. Subjects who understood FI exhibited more FI behavior than those who did not.

	(1) FI behavior	(2) FI behavior	(3) FI behavior
Verbal report in period (vr_{it})			−0.051
			(0.060)
No. of times partner chose In (OBS_{it})	0.018**	0.014	0.018****
	(0.009)	(0.010)	(0.003)
YES report	0.356****	0.135	0.194***
	(0.062)	(0.121)	(0.063)
MAYBE report	0.195***	0.103	0.161**
	(0.063)	(0.106)	(0.076)
NO report	−0.108	0.019	0.078
	(0.076)	(0.119)	(0.067)
No. of YES reports		0.113**	0.080****
		(0.056)	(0.011)
No. of MAYBE reports		0.051	0.021
		(0.051)	(0.013)
No. of NO reports		−0.079	−0.110****
		(0.062)	(0.013)
Treatment fixed effects	Yes	Yes	Yes
Constant	Yes	Yes	Yes
Observations	446	446	2648

Marginal effects of logit regressions. Bootstrapped standard errors in parentheses.
** $p < 0.05$.
*** $p < 0.01$.
**** $p < 0.001$.

we find that a larger fraction of participants understood FI, in the sense of providing at least one YES report, in T1 and T2 than in T4 ($P < 0.05$; Fisher's exact test). If we account for reports of players choosing *Out*, we find that 36% of participants in T4 showed an understanding of FI. This proportion is not significantly different from the 47% estimated in T1 and T2 ($P = 0.208$; Fisher's exact test). We summarize these findings as follows:

Result 2. In a given period, a small fraction of participants understood FI but lacked confidence to act upon this understanding.

4.4. Thinking and behavior

In rounds when reports were provided, a YES report was associated with FI-like behavior 69%, a MAYBE report 50%, and a NO report 16% of the time. These proportions are significantly different ($P < 0.01$ for all pairwise comparisons; Fisher's exact test). To quantify the effect of providing each kind of report on behavior, we add YES, MAYBE, and NO report dummies to the model specified by Eq. (1), restricting the data to periods in which reports were provided.[23] The results, collected in the first column of Table 5, show that a YES report increased the likelihood of being a FI player by about 36%, while a MAYBE report increased it by about 19%. Crucially, the marginal effect of a YES report is significantly greater than that of a MAYBE report ($P < 0.01$). We highlight this below:

Result 3. A significant fraction of FI behavior was driven by FI and not first mover advantage or other alternative ways of reasoning about the game.

We also find that subjects who produced at least one YES report exhibited FI-like behavior in more periods of the game *without reports* than subjects who did not ($P < 0.001$; Fisher's exact test). This holds if we restrict our attention to participants who did not provide any reports in the NO category ($P < 0.001$; Fisher's exact test), and if we restrict the sample to the first two periods of the experiment, i.e. to periods before any report was provided ($P < 0.001$; Fisher's exact test).

These results suggest that participants' behavior was driven by two factors: underlying understanding of FI and confidence in FI in a given period. To investigate this hypothesis, we control for the numbers of YES, MAYBE, and NO reports provided by the participant in the course of the experiment in addition to the rating of the report provided in a given period with belief elicitation. The estimated marginal effects are shown in the second column of Table 5. We find that the marginal effects of both YES and MAYBE dummies lose significance when reports in the other periods are controlled for ($P = 0.262$ and $P = 0.330$, respectively), which suggests a significant role of overall understanding of FI on determining participants' behavior. When we estimate the same model plus a belief elicitation dummy (vr_{it}) for every period of the experiment, we find that the marginal effect of vr_{it} is not significant ($P = 0.391$). Thus, the effect of answering questions on behavior was not uniform for all participants. In particular, the effect was significantly related to participants' understanding of FI.

[23] Since some subjects failed to provide reports, the report dummies are not collinear. The no report case serves as a baseline.

Table 6
The interaction of belief elicitation with underlying understanding of FI. Behavior is modeled as a function of belief elicitation dummy (vr_{it}), the number of YES reports, the number of MAYBE reports, the number of NO reports, and interactions between vr_{it} and each of the other variables. The first column of the table shows marginal effects of the vr_{it} variable in the cases where 0, 1, 2, or 3 YES reports are provided. The second column does the same for MAYBE reports, and the third column does the same for NO reports.

	(1) FI behavior	(2) FI behavior	(3) FI behavior
0	0.042	0.054	0.088**
	(0.030)	(0.036)	(0.035)
1	0.100****	0.089****	0.087****
	(0.022)	(0.020)	(0.024)
2	0.160****	0.125***	0.076*
	(0.040)	(0.041)	(0.042)
3	0.207***	0.162**	0.059
	(0.065)	(0.073)	(0.050)
Observations	2648	2648	2648

Marginal effects of a logit regression.
Bootstrapped standard errors in parentheses.
* $p < 0.10$.
** $p < 0.05$.
*** $p < 0.01$.
**** $p < 0.001$.

To explore this further, we estimate a model in which a participant's behavior is a function of a belief elicitation dummy (vr_{it}), the number of YES reports, the number of MAYBE reports, the number of NO reports, and interactions between vr_{it} and each of the other variables. Table 6 shows the marginal effects of belief elicitation as a function of how many reports of each kind the participant provided. The results show that the marginal effect of belief elicitation is not significant both for players who provided no YES reports and for those who provided three NO reports. We highlight this below as follows:

Result 4. Belief elicitation had a stronger effect on behavior of participants who understood FI.

Recall that the effect of belief elicitation on behavior was transient, and that some subjects produced FI reports in earlier periods and not in later ones. Together with Result 4 above, these findings suggest that belief elicitation increased subjects' confidence of FI rather than guided them toward an understanding of the FI logic.

5. Discussion

As noted by Blanco et al. (2010), an experiment that elicits participants' beliefs about their partners' behaviors is subject to a potential hedging problem. In our design, a subject was paid for guessing his partner's behavior only in the case the subgame is played. In this case, first movers played a risk-neutral best response to their beliefs 82% and second movers 83% of the time.[24] We conclude that most of the reports show no evidence of hedging.

Some studies, in contexts other than FI, have addressed the question of how participants' reported thinking relates to behavior. Most of these studies (using quadratic scoring rules) have shown that beliefs play a significant role in determining subjects' actions. This is the conclusion of Nyarko and Schotter (2002), Danz et al. (2012), and Hyndman et al. (2012), for instance. One prominent exception is Costa-Gomes and Weizsäcker (2008), who conclude that "stated beliefs reveal deeper strategic thinking than [. . .] actions. On average, [subjects] fail to best respond to their own stated beliefs in almost half of the games." Unlike Costa-Gomes and Weizsäcker (2008), and as highlighted in the paragraph above, we find that subjects in our experiment tended to best respond to their stated beliefs. Moreover, we find strategic sophistication to be highly predictive of behavior (Result 3).

Other papers have looked at the effect of eliciting incentivized and non-incentivized beliefs on behavior. While the results are mixed, several studies, such as Croson (2000), have found that incentivized belief elicitation has a significant effect on subjects' choices. Ours also finds this effect to be significant. Schotter and Trevino (2014) interpret the existing evidence to suggest that to the extent belief elicitation alters behavior, it makes stable, best response behavior emerge sooner. Our results are in part consistent with this interpretation: The prevalence of FI-like behavior grows over time even in absence of belief elicitation, but asking participants questions makes such behavior more likely. We also, however, observe *transient* effects of belief elicitation on choice. This suggests an interesting additional feature in the learning process identified in Schotter and Trevino (2014).

To the extent that our analysis above relies on the evaluators' ratings of strategic thinking, it is important to assess the quality of these ratings. To this end, we compared the evaluators' ratings to our own in the longer treatments of the

[24] We use the answer to F1 to quantify the first mover's beliefs about the second mover's behavior in the subgame, and the answer to S8 to quantify the second mover's beliefs about the first mover's behavior in the subgame.

experiment (T1 and T2). We found a high degree of correlation between our ratings of first movers' reports and those provided by the evaluators ($r = 0.86$, $P < 0.001$). While this high correlation is not surprising, given that we instructed the evaluators how to rate the reports, it serves as a good test of the evaluators' performance. Note that there is little reason to be concerned about experimenter demand effects. First, we take our definitions and examples of what constitutes forward induction thinking to be uncontroversial. Second, the evaluators were given no reason to believe that we, the experimenters, were interested in finding that many participants exhibit forward induction reasoning, or that few participants do. Indeed, what we told the evaluators is that we are interested in their true beliefs (see footnote 11).

6. Conclusion

We find strong evidence that FI guides thinking and behavior of a substantial number of experimental participants. The display of this ability is in part influenced by the activity of analyzing and explaining one's thought process, and in part by limited observation of the actions of others: The more often a subject observes his partners choose *In*, the more likely he is to exhibit FI thinking. There is also a link between understanding FI and behaving according to its logic: Subjects that show evidence of FI in their reports are more likely to exhibit FI behavior. Thus, FI, and not other factors, such as first mover advantage, induces the largest frequencies of the behavior predicted by theory.

Previous experimental research had provided only mixed evidence in support of the FI hypothesis. Our results, based on a different method which provides richer information suggest instead that FI thinking models accurately the reasoning process of a significant amount of participants. This confirms existing findings on the specifically human ability to make predictions by pure reasoning (Teglas et al., 2011). In the game we used, the second mover observes information on their opponent's first stage behavior, and, guided by an abstract assumption on players' behavior (FI), makes a prediction about a novel situation (the second stage of the game). It is likely that previous experimental studies of FI did not provide supportive evidence because FI is not easily identified by behavior in a single game: Using only behavioral data, it is impossible to quantify how much the effect of the focal point contributes to ((*In*, *Top*), *Left*) being played. Our methods circumvent this problem by treating FI as a cognitive process and looking for it in participants' explanations of their thoughts and behavior, rather than their selected actions.

Finally, our subjects were inexperienced, and were incentivized with amounts typical of laboratory economics experiments. It's possible that experienced or professional players in real life environments understand and use the forward induction logic more frequently.[25] This hypothesis should be tested in future research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jebo.2016.05.002.

References

Armantier, O., Treich, N., 2013. Eliciting beliefs: proper scoring rules, incentives, stakes and hedging. Eur. Econ. Rev. 62, 17–40.
Bagwell, K., Ramey, G., 1996. Capacity, entry, and forward induction. RAND J. Econ. 27, 660–680.
Battigalli, P., Friedenberg, A., 2012. Forward induction reasoning revisited. Theor. Econ. 7, 57–98.
Ben-Porath, E., Dekel, E., 1992. Signaling future actions and the potential for sacrifice. J. Econ. Theory 57, 36–51.
Blanco, M., Engelmann, D., Koch, A.K., Normann, H.-T., 2010. Belief elicitation in experiments: is there a hedging problem? Exp. Econ. 13, 412–438.
Brandts, J., Cabrales, A., Charness, G., 2007. Forward induction and entry deterrence: an experiment. Econ. Theory 33, 183–209.
Charness, G., Dufwenberg, M., 2006. Promises and partnership. Econometrica 74, 1579–1601.
Cooper, R., DeJong, D., Forsythe, R., Ross, T., 1993. Forward induction in the battle-of-the-sexes games. Am. Econ. Rev. 83, 1303–1316.
Costa-Gomes, M.A., Weizsäcker, G., 2008. Stated beliefs and play in normal-form games. Rev. Econ. Stud. 75, 729–762.
Croson, R.T., 2000. Thinking like a game theorist: factors affecting the frequency of equilibrium play. J. Econ. Behav. Organ. 41, 299–314.
Danz, D.N., Fehr, D., Kübler, D., 2012. Information and beliefs in a repeated normal-form game. Exp. Econ. 15, 622–640.
Fischbacher, U., 2007. z-Tree: Zurich toolbox for ready-made economic experiments. Exp. Econ. 10, 171–178.
Frechette, G., 2015. Laboratory experiments: Professionals versus students. In: Frechette, G., Schotter, A. (Eds.), Handbook of Experimental Economic Methodology. Oxford University Press, pp. 360–390.
Govindan, S., Wilson, R., 2009. On forward induction. Econometrica 77, 1–28.
Huck, S., Müller, W., 2005. Burning money and (pseudo) first-mover advantages: an experimental study on forward induction. Games Econ. Behav. 51, 109–127.
Hyndman, K., Ozbay, E.Y., Schotter, A., Ehrblatt, W.Z., 2012. Convergence: an experimental study of teaching and learning in repeated games. J. Eur. Econ. Assoc. 10, 573–604.
Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus and Giroux, New York.
Kohlberg, E., Mertens, J.-F., 1986. On the strategic stability of equilibria. Econometrica 54, 1003–1037.
Man, P.T., 2012. Forward induction equilibrium. Games Econ. Behav. 75, 265–276.
Müller, C., 2014. Robust Virtual Implementation Under Common Strong Belief in Rationality, Mimeo.
Nyarko, Y., Schotter, A., 2002. An experimental study of belief learning using elicited beliefs. Econometrica 70, 971–1005.
Ponssard, J.-P., 1991. Forward induction and sunk costs give average cost pricing. Games Econ. Behav. 3, 221–236.
Rick, S., Weber, R.A., 2010. Meaningful learning and transfer of learning in games played repeatedly without feedback. Games Econ. Behav. 68, 716–730.
Samuelson, L., 2005. Economic theory and experimental economics. J. Econ. Lit. 43, 65–107.

[25] See Frechette (2015) for a review of laboratory experiments in economics studying the differences between professional and subject populations.

Schotter, A., Trevino, I., 2014. Belief elicitation in the laboratory. Annu. Rev. Econ. 6, 103–128.
Selten, R., Sadrieh, A., Abbink, K., 1999. Money does not induce risk neutral behavior, but binary lotteries do even worse. Theory Decis. 46, 213–252.
Sonnemans, J., Offerman, T.T., 2001. Is the Quadratic Scoring Rule Behaviorally Incentive Compatible? Mimeo.
Teglas, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J.B., Bonatti, L.L., 2011. Pure reasoning in 12-month-old infants as probabilistic inference. Science 332, 1054–1059.
Xiao, E., Houser, D., 2005. Emotion expression in human punishment behavior. In: Proc. Natl. Acad. Sci. U. S. A, pp. 7398–7401.