King of the Hill: Giving Backward Induction its Best Shot*

Martin Dufwenberg Matt Van Essen

March 12, 2017

Abstract

We study a class of deceptively similar games, which however have different player sets and predictions that vary with their cardinal**ity.** The game-theoretic principles involved are compelling as predictions rely on weaker and less controversial epistemic foundations than needed to justify backward inductions more generally. Is the account empirically relevant? We design and report results from a relevant experiment.

KEYWORDS: backward induction, interactive epistemology, player set cardinality, experiment

JEL codes: C72, C92

1 Introduction

We offer two independent, and arguably equally important, motivations:

Motivation #1

Some classes of games can be meaningfully parameterized by the cardinality of the player set (N), and shown to possess properties that depend in interesting ways on N. For example, the class of N-player Cournot games nicely links the cases of monopoly ($N = 1$) and perfect competition ($N \to \infty$).

*MD: University of Arizona, University of Gothenburg, and CESifo, martind@eller.arizona.edu; MVE: University of Alabama, mjvanessen@cba.ua.edu. Part of the research was done while MD was on faculty at Bocconi University. We are grateful to that institution for providing funds for the experiment. We thank Geir Asheim, Pierpaolo Battigalli, Peter Norman Sørensen, and the participants at several seminar and conference presentations for helpful comments and discussion.

We explore a class of N-player games where predictions systematically vary with N in a different and intriguing way. The following problem illustrates:

> Consider N "subjects" in a line, in front of a "king" on a throne. The subject first-in-line must choose whether or not to dethrone the king. If not, the game ends (and all subjects go home). If the subject dethrones the king then he becomes the new king. The subject next-in-line must now choose whether or not to dethrone the new king. If not, the game ends. If the subject dethrones the king then the subject becomes the new king, and the subject next-in-line must choose whether to dethrone, etc. The interaction continues until some subject does not detrone the sitting king, or until there is no subject in line. The most preferred outcome is to become a king who is not dethroned. Second best is to remain a subject. The worst outcome is to be dethroned. Will the original king be dethroned?

[*Stop and think before reading on!*]

This is an old problem which, however, seems little-known. Brams & Kilgour (1993; see footnote 5) describe one version, and a Google search reaches others (often with the players being lions & lambs instead of subjects & kings). One of us learned about it from Jacob Goeree twenty years ago. Casual empiricism (try it on friends & colleagues!) suggests most people never heard of it, and find it hard to see through the thicket. However, reasoning by backward induction (BI) one realizes that the solution exhibits an odd-even effect. The original king will be dethroned if N is odd, not dethroned if N is even.

The economic relevance should be clear. The king could be a warlord or dictator, and the subjects potential competitors (foot-soldiers or ministers), and we get examples concerning geopolitical stability. Alternatively, consider voting procedures, where N parties or individuals sequentially reject and propose budgets.[1] We conjecture that parallel problems may also arise in societies with weak property rights (cf. Kaplow & Shavell 1996, Bar-Gill & Persico 2016), where agents may take each others' goods. Odd-even effects arise also in behavioral models of intertemporal choice, e.g. $\beta\delta$- models of procrastination (O'Donoghue & Rabin

[1] See Stewart (1999) for a hilarious related analysis: N pirates sequentially propose how to divide their loot, followed by voting whether to accept the proposal or throw the proposer overboard. Conclusions do not exhibit an odd-even effect, but depend starkly on N.

1999).[2]

TTBOOK, no one has explored the empirical relevance of odd-even effects. Using a variety of king of the hill (or KOH) games, that match versions of the story we told above, we design a series of lab-experiments to tackle this task.

Motivation #2

Among scholars who worked on the epistemic foundations of game-theoretic solution concepts, BI (in extensive games of perfect information) is a controversial procedure which it takes strong and questionable assumptions to justify. In a recent contribution, Arieli & Aumann (2015, p. 460) argue that "backward induction reasoning applies only to simple games," by which they mean games where each player moves just once. Let us elucidate. When assessing the plausibility of BI, there are two problems:

First, Arieli & Aumann favor an epistemic condition launched by Battigalli & Siniscalchi (2002) and called "rationality and common strong belief in rationality" (RCSBR). It characterizes Pearce's (1984) classic notion of extensive form rationalizability, and the BI path is implied, but examples can be constructed such that the BI solution (including off-path choices) is not.[3] Hence players do not necessarily reason according to BI. In many games where players move once, however, RCSBR implies the BI solution.[4]

Also for many games outside that "simple" class RCSBR implies the BI solution. Yet, and this is the second problem with BI, another objection can then be raised. First articulated by Kaushik Basu and Phil Reny in the mid-1980s, it goes something like this:[5] Suppose player i deviates from the BI path, and j is

[2] One can show that if (say) Ann must select one of N consecutive days on which to do a boring task, then, for appropriate parameters reflecting her inclination to instant gratification ($\beta < \delta$) and awareness of this ("sophistication"), applying BI regarding the choices of her future selves, she will do the task immediately iff N is odd. We thank Geir Asheim for this example.

[3] See Figure 2 in Battigalli & Siniscalchi and Figure 1 in Arieli & Aumann and the surrounding text. Key credit in this connection to Reny (1992) who explored closely related themes and inspired much leading up to RCSBR (including the poster game; see his Figure 3).

[4] Each player moving once is not a sufficient condition. For example, a player who moves after another player made a strictly dominated (irrational) choice may assume (under RCSBR) that other players are irrational as well, and so not follow BI. In such a case it would, however, seem that other arguably appealing epistemic assumptions imply BI. Assume that the observed behavior by co-player j does not effect i's beliefs about $k \neq j$, coupled with initial common belief in rationality.

[5] Examples of references that embrace versions of this line of thinking include Basu (1988), Reny (1988, 1993), Binmore (1987), Ben-Porath (1997), Gul (1997), and Asheim & Dufwenberg (2003) to whom we refer for more commentary and a model which shows how other than RCSBR, but arguably attractive, epistemic assumptions ("common certain belief of full admissible consistency") admit play to leave the BI path.

asked to move and has to take into account that i will move again. BI, implicitly, calls for j to assume that i will conform with BI in the future. Maintaining that belief is awkward, since j has seen evidence that i is, in fact, not making choices consistent with BI. If j therefore entertains the possibility that i may not conform with BI going forwards, he may have reason to deviate from BI himself.[6] But if this is true, i may have an incentive to deviate from the BI path to start with! The power of this argument is seen most starkly in centipede games (Rosenthal 1981) or chain store paradox games (Selten 1978). To overcome it, and deduce that players will behave according to BI, scholars have to make "strong assumptions about the players' belief-revision policies." That quote, from Battigalli & Siniscalchi (p. 374), refers to RCSBR. A similar remark would be appropriate for other epistemic conditions that imply BI.[7]

The literature on the empirical relevance of BI is largely centered on the centipede game, a context where each player moves multiple times and Basu-Reny objections applies.[8] No one has explored the empirical relevance of BI using games where the problems are irrelevant. Using KOH games, which have the property that RCSBR implies BI, we design a series of lab-experiments to tackle this task. Basu-Reny objections to BI have no bite; if i deviates from the BI path, this offers no presumption regarding subsequent play as i has no further choice.

More...

Apart from the two motivations already described, we are also interested in aspects of experience and insight. Since even colleagues who know game theory stumbled when we posed the problem to them, we conjecture that this happens because if N is a big number many fail to realize that they can apply backward induction. In that case, perhaps performance is enhanced if before considering a longer game (with a higher N) subjects may play and experience a shorter game (with a lower N).[9] We explore experimental treatments reflecting that idea.

We furthermore use two different versions of KOH games that differ regarding whether subjects move in sequence (as in the above problem) or simultaneously (as

[6] Up to here, the point was (essentially) made already by Luce & Raiffa (1957, pp. 80-81).

[7] For more on such other-than-RCSBR epistemics, see Asheim (2002) or Perea (2014).

[8] The BI solution for selfish players then does not predict particularly well. See e.g. McKelvey, & Palfrey (1992), Fey, McKelvey & Palfrey (1996), Rapoport, Stein, Parco & Nicholas (2003), Bornstein, Kugler & Ziegelmeyer (2004), and Levitt, List & Sadoff (2011). See also Binmore, McCarthy, Ponti, Samuelson & Shaked (2001) who report non-support for BI, in a study that does not employ centipede games, and yet many key comparisons involve players moving multiple times (note e.g. the results mentioned at the top of p. 85).

[9] Dufwenberg, Sundaram & Butler (2010) explore a similar issue in an otherwise different game (where the issue concerns the epiphany that one may have a dominant strategy).

if they surrounded the king). We call these versions the "line game" and the "ring game." Only the former has perfect information, but a BI argument (supported by RCSBR) nevertheless applies also to the ring game.[10] This allows us to explore the robustness of BI to concerns that some players may make a mistake.

Section 2 presents, and theoretically explores, all versions of our KOH games. Section 3 contains everything related to the experiment. Section 4 concludes.

2 King of the Hill Games

We study several versions of two kinds of sequential "capture" games. There are always three player roles: (1) King of the Hill; (2) Subject; and (3) Dethroned King. At the beginning of a game, everyone is put in the subject role, but may attempt to become king by charging the hill. A player's payoff depends on the role he finds himself in at end of the game:

1. If he is *King of the Hill* he receives a payoff of 8.
2. If he is a *Subject* he receives a payoff of 4.
3. If he is a *Dethroned King* he receives a payoff of 0.

Our two games differ in how subjects may charge the hill. However, they yield comparable predictions. We now describe the rules for each game.

The Line Game

The line game takes place over rounds. Subjects will be numbered 1 through N. This number determines the round in which a subject will make their choice. Subject 1 gets to make his decision in Round 1; Subject 2 gets to make his decision in Round 2, etc. In each round, the subject whose turn it is must decide whether to "Charge the Hill" or to "Stay Idle." These choices have the following results. If the subject chooses to "Stay Idle," then the game is over. If, however, the subject chooses to "Charge the Hill," then he becomes King of the Hill and the game continues to the next round. If there was a King of the Hill from a previous round

[10] This is in analogy to how a finitely repeated prisoners' dilemma (FRPD) can be solved by BI, despite the stage game having simultaneous moves. BI paradoxes comparable to those for centipede games have been discussed for the FRDP. See Petitt & Sugden (1989) for an early contribution, and Asheim & Dufwenberg (2003, section 4.3) for more.

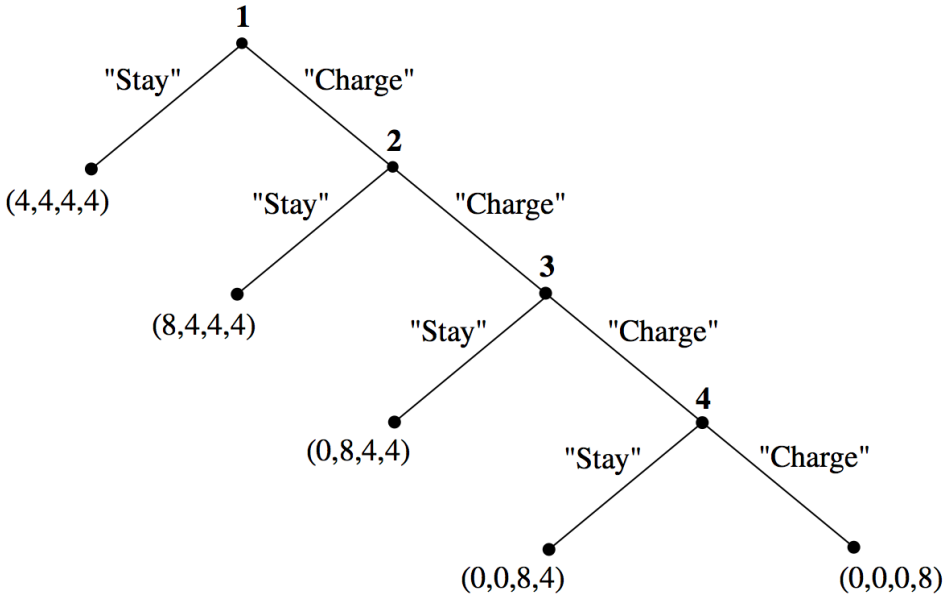

Figure 1: Line Game ($N = 4$)

and a subject chose to charge the hill, then the King of the Hill of the previous round becomes a Dethroned King.

The game continues until either there are no more subjects left to "Charge the Hill" or we reach a round where a subject decides to "Stay Idle." The maximum number of rounds is N.

Figure 1 illustrates the extensive form for the line game with four players.

The Ring Game

The ring game takes place over rounds. In every round, *each* subject must decide whether to "Charge the Hill" or to "Stay Idle." If no subject chooses to "Charge the Hill," then the game is over. If at least one subject chooses to "Charge the Hill," then a new King of the Hill is determined by randomly selecting one of the subjects who chose to "Charge the Hill." If there was a King of the Hill from a previous round, and if someone charged the hill, then the King of the Hill of the previous round now becomes a Dethroned King. Once a player has been made King of the Hill, that player makes no more decisions for the rest of the game regardless of whether he is currently King of the Hill or a Dethroned King.

The game continues until either there are no more subjects left to "Charge the Hill" or we reach a round where all subjects decide to "Stay Idle." The maximum number of rounds is N.

Theoretical Prediction

The backward induction (BI) prediction for both games is as follows. Clearly, in the last round, there is only one subject who can "Charge the Hill" and this subject always should. In the second to last round, a subject does not want to become king since he will be surely dethroned in the subsequent stage so all subjects should choose to stay idle in this round. This of course would end the game. As a result, subjects in the second to last round should all charge the hill since the game will be over in the subsequent round. This pattern continues. The BI prediction for the two games therefore yields the same pattern of behavior for each fixed group size. The following table summarizes the theoretical suggestion for $N = 2$, 3, and 4, but the pattern for higher N should be clear.

Game	N	Theory
Ring or Line	2	S, C
	3	C, S, C
	4	S, C, S, C

.

Thus, when N is even we expect both games to end after the first round with all players finishing the game as subjects. When N is odd we expect both games to end after the second round with a single king and $N-1$ subjects. The BI outcome (= path) of the game depends only on whether the number of players is odd or even. Figure 2 illustrates the BI solution for the line game with four players.

An important distinction between the line and ring games is that only in the former do each player possess a single information set, so one may wonder whether the second objection to BI that we described in the introduction apply to the ring game. They do not, for a subtle reason. Namely, we consider a version of the ring game such that after each round players are told who became the new king, but not what choice any particular co-player just made (except the one made king, of course, but that player makes no more move anyway).

Robustness

So far we have determined a unique BI prediction for each KOH game. We now explore the robustness of the prediction with respect to others' adherence to

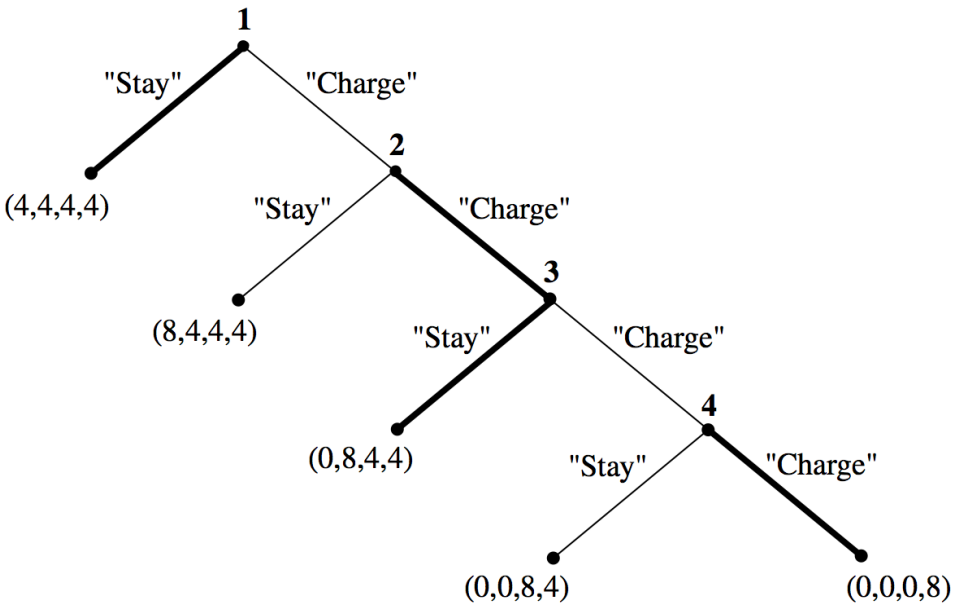

Figure 2: BI Prediction Line Game ($N = 4$)

the theory. This is done by considering the decision of a rational player i whose opponents, in each round, each choose the BI predicted action with probability $p \in [\frac{1}{2}, 1]$, where p is independent and identical in each round.

We ask two questions in the context of this model: First, "For what values of p is the BI prediction still optimal for i?" These values are said to be "*BI consistent.*" Identifying this set provides a measure of robustness for the BI prediction. Second, "Is one of the KOH games more robust than the other?" Despite the similarities between the games it seems intuitive that the ring game should be more sensitive to changes in p. A player in the three-player ring game, for instance, will "Charge" in round 1 only if he is relatively sure that the other two players will choose "Stay" in round 2. In contrast, a player in the three-player line game will "Charge" in round 1 only if he is relatively sure that the player in the second position will choose "Stay" in round 2. In short, since more people are making a decision in each round in the ring game deviations from the prediction are compounded.

We now compute and compare the BI consistent p for the two games. The differences we observe generate several testable hypotheses for the experiment.

In the line KOH game, the BI consistent parameters are easy to identify. In the last round, player i should always charge for all p. Consider round $N-k$, where k is odd. If player i stays, then he gets 4 for certain. If i charges, then he gets $8(1-p)$. So, p is consistent if $p \geq \frac{1}{2}$. Now suppose k is even. If player i stays, then he gets 4 for certain. If i charges in then he gets $8p$. Thus, p is consistent if $p \geq \frac{1}{2}$. We summarize as follows:

Observation 1: *In the line game, all $p \in [\frac{1}{2}, 1]$ are BI consistent.*

Now consider the ring KOH game. In contrast to the line KOH game, the BI consistent parameters for the ring KOH are more difficult to identify. We therefore characterize this set by identifying the payoff v_{N-k} associated with the optimal action for player i in each round $N-k$ given p, for each $k = 0, ..., N-1$.

In the last round ($k = 0$), it is clearly optimal to for player i to charge – independent of p. The payoff associated with reaching round N is $v_N(p) = 8$.

Next, we compute $v_{N-k}(p)$. In round $N-k$, there are $k+1$ players remaining in the role of subjects. We first suppose that k is an odd number. If player i chooses to charge, then $m = 0, 1, ..., k$ of the other k players choose to charge with probability

$$\binom{k}{m}(1-p)^m p^{k-m}.$$

In this case, he becomes king with probability $\frac{1}{1+m}$ and remains a subject with probability $\frac{m}{1+m}$. If he becomes king, then play goes to round $N-k+1$ where he expects $8\left(1-p\right)^k$. If he does not become king, then he remains a subject, play goes to round $N-k+1$ and he gets an expected payoff of $v_{N-k+1}(p)$. Thus, the expected payoff of charging in round $N-k$ is

$$\sum_{m=0}^{k}\binom{k}{m}(1-p)^m p^{k-m}\left(\frac{8(1-p)^k}{1+m}+\frac{m}{1+m}v_{N-k+1}(p)\right)$$

The expected payoff of choosing stay is $4p^k+(1-p^k)v_{N-k+1}(p)$. The optimal action gives i the larger payoff. Thus,

$$v_{N-k}(p)=\max\left\{\begin{array}{c}4p^k+(1-p^k)v_{N-k+1}(p),\\ \sum_{m=0}^{k}\binom{k}{m}(1-p)^m p^{k-m}\left(\frac{8(1-p)^k}{1+m}+\frac{m}{1+m}v_{N-k+1}(p)\right)\end{array}\right\}.$$

Now suppose k is even. If player i chooses to charge, then $m=0,1,...,k$ of the other k players choose to charge with probability

$$\binom{k}{m}p^m\left(1-p\right)^{k-m}.$$

In this case, he becomes king with probability $\frac{1}{1+m}$ and remains a subject with probability $\frac{m}{1+m}$.

If he becomes king, then play goes to round $N-k+1$ where he expects $8p^k$. If he does not become king, then he remains a subject, play goes to round $N-k+1$ and he gets an expected payoff of $v_{N-k+1}(p)$. Thus, the expected payoff of charging in round $N-k$ is

$$\sum_{m=0}^{k}\binom{k}{m}p^m\left(1-p\right)^{k-m}\left(\frac{8p^k}{1+m}+\frac{m}{1+m}v_{N-k+1}(p)\right).$$

The expected payoff of choosing stay is

$$4(1-p)^k+(1-(1-p)^k)v_{N-k+1}(p).$$

The optimal payoff for the round is therefore

$$v_{N-k}(p)=\max\left\{\begin{array}{c}4(1-p)^k+(1-(1-p)^k)v_{N-k+1}(p),\\ \sum_{m=0}^{k}\binom{k}{m}p^m\left(1-p\right)^{k-m}\left(\frac{8p^k}{1+m}+\frac{m}{1+m}v_{N-k+1}(p)\right)\end{array}\right\}.$$

Finally, in order to for p to be BI consistent, the payoff associated with the BI prediction for each round must be larger than the other choice. The following result is immediate.

Observation 2: *In the ring game, p is BI consistent if and only if for $k = 1,...,N-1$ we have*

$$4p^k + (1-p^k)v_{N-k+1}(p) \geq \sum_{m=0}^{k} \binom{k}{m} (1-p)^m p^{k-m} \left(\frac{8(1-p)^k}{1+m} + \frac{m}{1+m} v_{N-k+1}(p) \right)$$

when k is odd and

$$\sum_{m=0}^{k} \binom{k}{m} p^m (1-p)^{k-m} \left(\frac{8p^k}{1+m} + \frac{m}{1+m} v_{N-k+1}(p) \right) \geq 4(1-p)^k + (1-(1-p)^k) v_{N-k+1}(p)$$

when k is even.

The above result characterizes the p that are BI consistent. Given this characterization it is possible to compute the set of consistent beliefs numerically. This is done recursively starting with round $N-k$, for $k = 0, 1, ..., N-1$. The table below provides the sets of consistent beliefs for group sizes $N = 2$, 3, 4, and 5 for both the ring game and the line game.

N	Consistent p (Ring)	Consistent p (Line)
2	$\left[\frac{1}{2}, 1\right]$	$\left[\frac{1}{2}, 1\right]$
3	$[0.774\,2, 1]$	$\left[\frac{1}{2}, 1\right]$
4	$[0.774\,2, 1]$	$\left[\frac{1}{2}, 1\right]$
5	$[0.8608, 1]$	$\left[\frac{1}{2}, 1\right]$

Several things are noteworthy. For $N = 2$, *any belief* $p \in \left[\frac{1}{2}, 1\right]$ is BI consistent for either game. Things change, however, when there are more than two players. For $N = 3$, the set of consistent beliefs contracts sharply for the ring game and remains constant for the line game. This follows since, in the ring game, an additional player is added to the decision making process which increases the likelihood of being dethroned if a player charges in the first round. Players are therefore reluctant to charge unless p is sufficiently close to 1. In the line game, the decision problem remains the same. For $N = 4$, the set of consistent beliefs does not change from $N = 3$ to $N = 4$ in either game. For the ring game, the binding beliefs are the ones needed to support charging in round 2 which is the same as charging in round 1 in the $N = 3$ game. This pattern continues. The set

of consistent beliefs in the line game stay constant whereas the set contracts in the ring game at each odd numbered N (i.e., games where the BI prediction for the first round is to charge). Thus, for larger N the line game is more robust to changes in beliefs than the ring game as expected.

3 The Experiment

3.1 Predictions and Treatments

The experiment conducted involves the two KOH games. The primary hypothesis derived from standard theory is that the observed play will conform to the BI prediction in both games for all group sizes. As a consequence, we expect to see the type of odd-even effect described in Motivation #1. Cognitive limits and inexperience, however, suggest that this prediction may not always obtain. Thus, in the experiment, we have treatments designed to test the robustness of this prediction. In particular, the treatments included one-shot versions of each game where the number of players in the game are $2, 3$, and 4. The larger games are more complicated since they require a longer chain of reasoning to arrive at the backward induction solution.

We have seen that larger ring games also require more restrictive beliefs about the play of others to support that prediction. Based on these two observations we expect the number of failures of the prediction to be increasing in N. In addition, the earlier discussion of BI consistency also suggests that we may see more departures from that theory in the ring game than in the line game for the three- and four-player treatments.[11]

The choice of using a one-shot game (i.e., no practice rounds) is perhaps extreme. In particular, it does not give the players much of a chance to learn. In further treatments, we therefore explored the impact of experience by exposing players to a two-player game first and then let them play either the three-player or four-player versions on the same game. The two-player game is easy to solve and the prediction is robust to changes in p. Effectively, by exposing players to a subgame, we allow them to adjust their beliefs about play in later rounds. We expect less failures of the BI prediction in the "experienced" treatments.

[11] This is assuming a uniform prior about the probability parameter p in our model of BI consistency.

In summary, the experiment consisted of the following eight different treatments.

N	Ring game	Line game
3	x	x
4	x	x
2-then-3	x	x
2-then-4	x	x

We do not run separate treatments for the one-shot two player games. Since all subjects in the experience treatments always play the two player game first (i.e., the "2-then-3" and the "2-then-4" treatments), the two-player games are played under the same conditions as the one shot three- and four-player games. We therefore treat these two-player game observations as one-shot observations when making statistical comparisons. In the results section, we distinguish the three- and four-player games where subjects have experience by labeling them either Ring (E) or Line (E).

3.2 Procedures

All lab sessions were conducted at the University of Alabama. Subjects were undergraduate students recruited via E-mail from large section sociology and economics classes. Roughly twenty subjects participated in each session.

Upon arrival, subjects were checked-in and randomly assigned to a seat in the classroom where they were given a set of instructions for the treatment being run. Only one treatment was run in each session. These sessions typically lasted about 30 minutes and no subject participated in more than one treatment.[12]

In each session, subjects were asked to read the instructions to themselves and, subsequently, the experimenter would read the instructions aloud. Any questions that the subjects had were answered privately. The same experimenter was present at each of the sessions. After the instructions had been completed and all questions had been answered the game was started.

This experiment was completely paper based and proceeded as follows: First, subjects were directed by the instructions to read a short summary statement for each round and then decide what they would do if placed in that situation. Once all contingent decisions had been made, the experimenter collected the decision sheets

[12] The instructions are available in the appendix.

			Individual Strategy Choice Frequency												
			2 Player Game				3 Player Game					4 Player Game				
	Treatment	# of Observations	SC	CC	SS	CS	CSC	SSC	CCC	SSS	Other	SCSC	SSSC	CCCC	SSSS	Other
1	Ring 3	23					1	16	2	2	2					
2	Ring 4	19										0	9	5	1	4
3	Ring 2 & 3	24	20	1	2	1	10	12	0	0	2					
4	Ring 2 & 4	18	13	0	4	1						3	11	0	1	3
5	Line 3	21					5	9	0	5	2					
6	Line 4	21										4	7	0	1	9
7	Line 2 & 3	17	13	1	2	1	10	7	0	0	0					
8	Line 2 & 4	21	20	1	0	0						9	8	0	3	1
	Theory		X				X					X				

Figure 3: Summary Data

from the subjects and randomly matched people into groups and assigned player roles.[13] The game was then played out for each group according to the directions submitted by the players in that group. Random decisions were determined using a Bingo cage. Subjects were privately paid their total earnings at the end of the session. No exchange rate was used.

3.3 Results

Figure 3 presents summary data for all of the treatments conducted in the experiment.

For each treatment, the data in the figure is broken up into (a) treatment; (b) the number of subjects in each treatment; and (c), since a version of the strategy method was used to collect the data, we report each subject's list of action choices – i.e., the list specifying which action each subject said they would have taken at each different point in the game.

This is the data used for all results. We break these up into several parts, as follows:

[13] In treatments that involved two games the experimenter ensured that no player was matched with the same people twice. Subjects were informed of this at the beginning of the experiment.

First, for each game/treatment, we examine to what extent the full BI solution is played. Of course, if the BI prediction was fully supported that would imply the odd-even effect. Since we have a complete plan of action from each player this is simply a matter of looking at the proportion of subjects who chose the BI prediction.

Second, we hold the game played constant (i.e., line or ring), and look at the robustness of the theoretical prediction as the number of group members increases – i.e., going from $N = 2$ to $N = 3$ to $N = 4$. Specifically, for each game, we report how the proportion of the players who made the theoretical prediction changes as we move to larger group sizes. Recall that the theory says there should be no difference.

Third, we compare the experienced treatments to the non-experienced treatments. Does exposure to a subgame increase the frequency of the theoretical prediction?

Fourth and finally, we compare the line and ring games. Our model of BI consistent beliefs suggested that range of beliefs that support the BI prediction is smaller for the ring game when $N \geq 3$.

Results, Part 1: Does the BI-Solution Work?

We begin our analysis by comparing the observed experimental behavior to the BI prediction. Since a version of the strategy method was used during the experiment, the data for each treatment/game specifies the list of action choices given by each participant.

In the experiment, the experimenter randomly matched people into groups and then played out the KOH game using the moves specified in these lists. The matching was random. Therefore we omit the description of the games that were played out in the experiment and focus, instead, on the proportion of players who chose the BI predicted list of actions. Recall that the BI prediction for the two-, three-, and four-player games are, respectively, SC, CSC, and SCSC.

Our first table displays, for each treatment, the count of individuals who played the BI prediction, the count of individuals who chose otherwise, as well as the proportion of BI play observed.

	BI	Other	% BI
Ring 2	33	9	0.79
Ring 3	1	22	0.04
Ring 4	0	19	0.00
Ring 3 (E)	10	14	0.42
Ring 4 (E)	3	15	0.17

	BI	Other	% BI
Line 2	33	5	0.87
Line 3	5	16	0.24
Line 4	4	17	0.19
Line 3 (E)	10	7	0.59
Line 4 (E)	9	12	0.43

The table above indicates that the BI prediction does well in both of the two-player games. It is also plain that for both the line and ring games, the BI prediction does not do well for the games with more than two players (albeit to different degrees).[14]

Results, Part 2: Robustness of Prediction to Changes in N

We now investigate how the proportion of players choosing the BI prediction varies with N across treatments. The table suggests that for both the ring and the line games the proportion of BI play fell with the number of players. We now test whether these drops going from $N = 2$ to $N = 3$ to $N = 4$ are significant. We use Fisher's exact test making a series of pairwise comparisons.[15] The research hypothesis is that games with smaller N will have a higher proportion of players who played according to the theoretical prediction. The associated null hypothesis is that the proportion of players who chose the theoretical prediction does not vary with N. The one-sided p-values for each of these tests are reported in the table below. In the conclusion column we indicate by (**) or (*) whether the null can be rejected at the 5% or 10% level respectively.

	Comparison	p-value
(1)	Ring 2 vs. Ring 3	0.000**
(2)	Ring 3 vs. Ring 4	0.548
(3)	Ring 3 (E) vs. Ring 4 (E)	0.080*
(4)	Line 2 vs. Line 3	0.000**
(5)	Line 3 vs. Line 4	0.500
(6)	Line 3 (E) vs. Line 4 (E)	0.257

[14] Figure 3 indicates that the majority of players that deviated from the BI prediction did so in favor of SSC in the three-player game, or SSSC in the four-player game. This deviation is perhaps understandable. It is a conservative course of action that guarantees that a player will never be dethroned.

[15] See, for instance, Siegal & Castellan (1988, p. 103).

The proportion of BI prediction play is initially high but falls in treatments where the number of players have been increased. In the one-shot game, the change is only significant going from two- to three-player games. We strongly reject equality of proportions of the two- and three-player treatments for both games in favor of the one-sided alternative. However, the change from three- to four-player groups is less dramatic. We cannot reject equal proportions for either game when three- and four-player treatments for either game.

The directional result is strengthened by examining the experienced treatments for the ring game. In particular, comparing Ring 3 (E) and Ring 4 (E), we can reject equal proportions for the three- and four-player treatments in favor of a higher success rate in the three-player ring game.

Results, Part 3: Experience

In the experienced treatments, the proportion of BI prediction play increased relative to the non-experienced treatment. In the three-player games, the proportion of BI play increased from 0.05 to 0.42 in the ring games and from 0.31 to 0.59 in the line games. In the four-player games, the proportion of BI play increased from 0 to 0.17 in the ring games and from 0.24 to 0.43 in the line games.

We test these comparisons statistically using Fisher's exact test making a series of pairwise comparisons. The research hypothesis is that the experienced treatments will have a higher proportion of players who made the theoretical prediction. The associated null hypothesis is that the proportion of players who chose the theoretical prediction does not vary between the one-shot game and the exposure to a smaller subgame. The one-sided p-values for each of these tests are reported in the table below.

	Comparison	p-value
(1)	Ring 3 vs. Ring 3 (E)	0.003**
(2)	Ring 4 vs. Ring 4 (E)	0.105
(3)	Line 3 vs. Line 3 (E)	0.031**
(4)	Line 4 vs. Line 4 (E)	0.091*

In the three-player treatments, the difference between the one-shot and the experienced treatments were all statistically significant at the 5% level. In the four-player treatments, the difference was only significant for the line game at the 10% level. In both of the experienced treatments, players were first exposed to a two-player subgame. The data suggests that this exposure was more impactive for those who subsequently continued onto the three player game.

Results, Part 4: Game Comparisons

Finally, comparing the proportion of BI prediction play in the two games (i.e., line vs. ring) we see that the line game tends to dominate the ring game for comparable sized treatments. We test these comparisons statistically using Fisher's exact test making a series of pairwise comparisons. Based on our model of BI consistency, the research hypothesis is that the Line treatments will produce higher proportions of players who made the theoretical prediction. The associated null hypothesis is that, for each fixed group size, the proportion of players who chose the theoretical prediction does not vary between the two games (i.e., ring and line). The one-sided p-values for each of these tests are reported in the table below.

	Comparison	p-value
(1)	Ring 3 vs. Line 3	0.074*
(2)	Ring 4 vs. Line 4	0.0655*
(3)	Ring 3 (E) vs. Line 3 (E)	0.2222
(4)	Ring 4 (E) vs. Line 4 (E)	0.0768*

In all comparisons except the Ring 3 (E) vs. Line 3 (E) we reject the null hypothesis in favor the alternative hypothesis. Specifically, the line game yields a higher proportion of players who played in accordance with BI. Although the effect is not very strong, this is consistent with our model of BI consistency.

4 Concluding remarks

Suppose the prediction in an N-player game depends on whether N is even or odd. It is then a corollary that if the game gets an added player, to become an $N+1$ player game, its solution will change. Alexandre Dumas may have been on to related insights. Take it from *The Man with the Iron Mask*:

> The Queen gave birth to a son. But when the entire court greeted the news with cries of joy, when the King had shown the newborn to his people and the nobility, when he gaily sat down to celebrate this happy event, the queen alone in her chamber, was stricken with more contractions, and then she gave birth to a second boy...The King ran back to his wife's chamber. But this time his face was not merry; it

> expressed something like terror. Twin sons changed into bitterness the joy caused by the birth of a single son.

Would the king have been happier with triplets? And horrified yet again with quadruplets? Dumas does not say, but we imagine he might have been curious about our study. We have examined a class of games where backward induction (BI) predicts an odd-even effect, somewhat in a related spirit.

Some objections that can be raised against the plausibility of BI in other games (e.g. centipede games) have no bite in ours. A person looking out for epistemic conditions to deem attractive may therefore have been hopeful that our design would be supportive of BI, and so of the odd-even effect.

Instead, such a person may find our actual results surprising and disappointing. With the exception of the games with just two stages, for the most part subjects did not play according to BI. If given an opportunity to gain experience through playing a simpler/shorter game before a longer one, subjects choose consistent with BI slightly more often, but that tendency is not particularly strong. Subjects also rely on BI strategies slightly more often in games where fewer succeeding co-players have the option to bring them down (by not conforming with BI), but again that tendency is not particularly strong.

Our goal has been to test BI in a context where BI is non-controversial from a interactive epistemology point of view, and where such tests have not previously been performed. Our goal has not been to also explain in depth why BI fails (beyond exploring our particular experience, tree-length, and line-vs-ring considerations), when such were the data. However, understanding what cognitive processes other than BI may be relevant is an important topic. Our remaining remarks are meant to inspire future such work:

Weathered experimentalist may be less surprised by our findings than believers in solid epistemics. A large literature explores aspects of players' deductive reasoning about each other in a variety of games. See Camerer (2003, ch. 5) for a nice review which covers e.g. guessing/beauty-contest, Bertrand, travelers' dilemma, e-mail, dirty-faces, and betting games, and, as mentioned, centipedes. Most are simultaneous-move games that are (weakly or strictly) dominance-solvable, so focused on BI (the centipede game is of course one exception). While the insights are therefore not directly comparable, it is still remarkable how Camerer's summary echoes our findings of limited inductive prowess: he offers (p. 202) that "the median number of steps of iterated dominance is two."

More recent related experimental literature increasingly relates to the level-k and cognitive hierarchy models: Level-k players best respond to some distribution of level-k' play, where $k' < k$, and level-0 follows some heuristic. See Costa-Gomes, Crawford & Iriberri (2013) for a survey.[16] Conclusions remain, by and large, analogous to Camerer's, op. cit. In a recent intriguing study Kneeland (2015) connects to the level-k scholarship (as well as to Bernheim's 1984 model of k-rationalizability), and argues that there is "possible misidentification due to the strong assumptions imposed" [especially: the specification of level-0 play, which impacts everything]. She develops an ingenious design (using a form of "ring game"), and reports "considerably more weight on higher-order types, R3-R4, than the level-k literature typically finds" (p. 2076). See her text for more details and motivation. She finds support for overall play being consistent with slightly more layers of what she calls "higher-order rationality" than we do (3-to-4 rather than 2). More research seems necessary to pin down why this is so, and how robust the patterns are. For now, we just note that subjects face different deductive reasoning tasks in Kneeland's and our design. In hers, subjects conduct iterated dominance calculations in simultaneous-move games; in ours they analyze a sequential-play games.

Where higher-order rationality fails, the need arises to develop new theories of strategic play. This is how the level-k/cognitive hierarchy literature came about, and this prompted Friedenberg, Kets & Kneeland (2016) to raise pertinent issues regarding whether it is "cognitive bounds" or limits to which people believe others exhibit (higher-order) rationality that shape behavior. The focus has been on simultaneous-move games. Developing extensions to sequential play may prove challenging, as aspects regarding how subjects perceive dynamic games must be tackled. Johnson, Camerer, Sen & Rymon (2002) provide exciting evidence regarding how far down a game-tree subjects actually look (using a technology where subjects' payoffs are hidden from view until clicked on, and experimenters observe subjects' clicks). Mantovani (2014) and Roomets (2010) take useful first steps towards modeling such considerations. However, that is early work and we propose that this major topic warrants much more attention.

[16] Important contributions getting this literature started include Stahl & Wilson (1994, 1995), Nagel (1995), Costa-Gomes, Crawford & Broseta (2001), and Camerer, Ho & Chong (2004).

5 Appendix: Instructions

We provide the instructions for the three player KOH Line game. The instructions for the other games are similar.

INSTRUCTIONS

This experiment concerns decision making in a "King of the Hill" game. If you read the instructions carefully and pay attention, you have the potential to earn some money. The decisions that are made may affect the payments of everyone else involved in the game. You will be randomly matched into groups of three.

In the game you may find yourself in each of three roles: (1) King of the Hill; (2) Subject; and (3) Dethroned King.

The game takes place over rounds. In the beginning of the game, everyone is a subject, but one of these subjects may attempt to become King of the Hill by charging the hill. More precisely, subjects will be in a line and they will be numbered 1 through 3. The number determines the round in which a subject may make a choice. Subject 1 gets to make his decision in Round 1; Subject 2 may get to make his decision in Round 2, etc. In each round, the subject whose turn it is must decide whether to "Charge the Hill" or to "Stay Idle."

These choices have the following results. If the subject chooses to "Stay Idle," then the game is over. If, however, the subject chooses to "Charge the Hill," then he becomes King of the Hill and the game continues to the next round.

If there was a King of the Hill from a previous round and a subject chose to charge the hill, then the King of the Hill of the previous round becomes a Dethroned King. The game continues, with the subject next in line choosing whether to "Charge the Hill" or to "Stay Idle," until either there are no more subjects left to "Charge the Hill" (that is, we have one King of the Hill and [two] Dethroned Kings) or we reach a round where a subject decides to "Stay Idle." The maximum number of rounds is [3].

How much money you make depends on which role you find yourself in at end of the game:

1. If you are King of the Hill when the game ends you receive $[8].
2. If you are a Dethroned King when the game ends you receive $[0].
3. If you are a Subject when the game ends you receive $[4].

You will now be asked to make several decisions.

Specifically, for each round, the experimenter will ask you to read a short summary statement and then make a choice. Note that we will not tell you right away whether you are subject 1, 2, or 3. Rather we will ask you what you would do in each of these cases. Once everyone has made a choice for each round, the summary sheets will be collected from your group. The experimenter will then randomly designate the members of your group as subject 1, 2, and 3 and use the appropriate choice for each subject to play out the King of the Hill game as directed by your group's decisions to determine an outcome for your group.

Your earnings from this outcome will be paid to you at the end of the experiment.

Are there any questions? Let's begin!

References

[1] Arieli, I. and Aumann, R. (2015), "The Logic of Backward Induction," *Journal of Economic Theory* 159, 443-464.

[2] Asheim, G. (2002), "On the Epistemic Foundation for Backward Induction," *Mathematical Social Sciences* 44, 121-144.

[3] Asheim, G. and Dufwenberg, M. (2003), "Deductive Reasoning in Extensive Form Games," *Economic Journal* 113, 305-325.

[4] Bar-Gill, O. and Persico, N. (2016), "Exchange Efficiency with Weak Ownership Rights," *American Economic Review*, forthcoming.

[5] Basu, K. (1988), "Strategic Irrationality in Extensive Games," *Mathematical Social Sciences* 15, 247-260.

[6] Battigalli, P. and Siniscalchi, M. (2002), "Strong Belief and Forward Induction Reasoning," *Journal of Economic Theory* 16, 356-391.

[7] Ben-Porath, E. (1997), "Rationality, Nash Equilibrium, and Backwards Induction in Perfect Information Games," *Review of Economic Studies* 64, 23-46.

[8] Bernheim, D. (1984), "Rationalizable Strategic Behavior, *Econometrica* 52, 1007-1028.

[9] Binmore, K. (1987), "Modeling Rational Players: Part I," *Economics and Philosophy* 3, 179-214.

[10] Binmore, K., McCarthy, J., Ponti, G., Samuelson, L. & Shaked, A. (2001), "A Backward Induction Experiment," *Journal of Economic Theory* 104, 48-88.

[11] Bornstein, G., Kugler, T. and Ziegelmeyer, A. (2004), "Individual and Group Decisions in the Centipede Game: Are Groups More 'Rational' Players?," *Journal of Experimental Social Psychology* 40, 599–605.

[12] Brams, S. and Kilgour, D. (1998), "Backward Induction is Not Robust: The Parity Problem and the Uncertainty Problem," *Theory and Decision* 45, 263-289.

[13] Camerer, C. (2003), *Behavioral Game Theory*, Russell Sage Foundation.

[14] Camerer, C., Ho, T.-H. and Chong, J.-K. (2004), "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics* 119, 861-898.

[15] Costa-Gomes, M., Crawford, V. and Broseta, B. (2001), "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica* 69, 1193-1235.

[16] Costa-Gomes, M., Crawford, V. and Irriberri, N. (2013), "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature* 51, 5-62.

[17] Dufwenberg, M., Sundaram, R., and Butler, D. (2010). "Epiphany in the Game of 21," *Journal of Economic Behavior & Organization* 75, 132-143.

[18] Fey, M., McKelvey, R. and Palfrey, T. (1996), "An Experimental Study of Constant-Sum Centipede Games," *International Journal of Game Theory* 25, 269–87.

[19] Friedenberg, A., Kets, W. and Kneeland, T. (2016) "Bounded Reasoning: Rationality or Cognition," unpublished manuscript, ASU, Northwestern, and UCL.

[20] Gul, F. (1997). "Rationality and Coherent Theories of Strategic Behavior," *Journal of Economic Theory* 70, 1-31.

[21] Johnson, E., Camerer, C., Sen, S. and Rymon, T. (2002), "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining," *Journal of Economic Theory* 104, 16-47.

[22] Kaplow, L. and Shavell, S. (1996), "Property Rules versus Liability Rules. An Economic Analysis." *Harvard Law Review* 109, 713-790.

[23] Kneeland, T. (2015), "Identifying Higher-Order Rationality," *Econometrica* 83, 2065-2079.

[24] Levitt, S., List, J., and Sadoff, S. (2011), "Checkmate: Exploring Backward Induction among Chess Players," *American Economic Review* 101, 975-990.

[25] Luce, D. and Raiffa, H. (1957), *Games and Decisions*, New York, Wiley.

[26] Mantovani, M. (2014), "Limited Backward Induction: Foresight and Behavior in Sequential Games," unpublished manuscript, University of Milan Bicocca.

[27] McKelvey, R. and Palfrey, T. (1992), "An Experimental Study of the Centipede Game," *Econometrica* 60, 803-36.

[28] Nagel, R. (1995), "Unraveling in Guessing Games: An Experimental Study," *American Economic Review* 85, 1313-1326.

[29] Nagel, R. and Tang, F. (1998), "Experimental Results on the Centipede Game in Normal Form: An Investigation on Learning," *Journal of Mathematical Psychology* 42, 356–84.

[30] O'Donoghue, T. and Rabin, M. (1999), "Doing it Now or Later," *American Economic Review* 89, 103-124.

[31] Pearce, D. (1984), "Rationalizable Strategic Behavior and the Problem of Perfection, *Econometrica* 52, 1029-1050.

[32] Perea, A. (2014), "Belief in the Opponents' Future Rationality," *Games and Economic Behavior* 83, 231-254.

[33] Petitt, P. and Sugden, R. (1989), "The Backward Induction Paradox," *Journal of Philosophy* 4, 169-182.

[34] Rapoport, A., Stein, W., Parco, J. and Nicholas, T. (2003), "Equilibrium Play and Adaptive Learning in a Three-Person Centipede Game," *Games and Economic Behavior* 43, 239–65.

[35] Reny, P. (1988). "Rationality, Common Knowledge and the Theory of Games," PhD Dissertation, Chapter 1, Department of Economics, Princeton University.

[36] Reny, P. (1992). "Backward Induction, Normal Form Perfection and Explicable Equilibria," *Econometrica* 60, 627-649.

[37] Reny, P. (1993). "Common Belief and the Theory of Games with Perfect Information," *Journal of Economic Theory* 59, 257-274.

[38] Roomets, A. (2010), "On Limited Foresight in Games," unpublished manuscript, University of Arizona.

[39] Rosenthal, R. (1981), "Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox," *Journal of Economic Theory* 25, 92-100.

[40] Selten, R. (1978), "The Chain Store Paradox," *Theory and Decision* 9, 127-159.

[41] Stahl, D. and Wilson, P. (1994), "Experimental Evidence on Player's Models of Other Players," *Journal of Economic Behavior and Organization* 25, 309-327.

[42] Stahl, D. and Wilson, P. (1995), "On Player's Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior* 10, 218-254.

[43] Stewart, I. (1999), "A Puzzle for Pirates," *Scientific American*, May issue, 98-99.