King of the Hill: Giving backward induction its best shot ☆

Martin Dufwenberg [a,b,c], Matt Van Essen [d]

[a] *University of Arizona, United States*
[b] *University of Gothenburg, Sweden*
[c] *CESifo, Germany*
[d] *University of Alabama, United States*

ARTICLE INFO

ABSTRACT

We study a class of deceptively similar games, which however have different player sets and backward induction (BI) predictions that vary with their cardinality. The game-theoretic principles involved are compelling as predictions rely on weaker and less controversial epistemic foundations than needed to justify BI more generally. Are the BI predictions empirically relevant for this class of games? We design and report results from a relevant experiment.

© 2018 Elsevier Inc. All rights reserved.

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 ($N \to \infty$).

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 dethrone 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 and Kilgour (1998; 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 20+ 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.

We believe there is applied potential. If the king were a sovereign state and the subjects hostile neighbors we might get an example concerning geopolitical stability. For another example, 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 and Shavell, 1996; Bar-Gill and 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 and Rabin, 1999).[2]

To the best of our knowledge, no one explored the empirical relevance of odd-even effects. Using a variety of king of the hill (KOH) games, that match versions of the story we told, 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 assumptions about the player's belief-revision policies" to justify.[3] A key objection, first articulated by Kaushik Basu and Phil Reny in the mid-1980s, goes something like this[4]:

Suppose player i deviates from the BI path, and j is 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.[5] 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) in which all players move multiple times, or chain store paradox games (Selten, 1978) where one player does so. The experimental literature on the empirical relevance of BI is largely centered on centipede games, and the BI solution for selfish players then does not predict particularly well.[6] Since Basu–Reny objections apply, it is natural to wonder whether BI would work better in games in which that were not the case.

We address this issue by designing a series of lab-experiments involving KOH games such that, in the basic versions, each player moves only once. Basu–Reny objections 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 main motivations already described, we are interested in exploring two additional issues each of which will be reflected using particular treatment variables. First, we consider two different versions of KOH games that differ regarding whether subjects move in sequence (as in the above problem) or simultaneously (as if they surrounded

---

[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.

[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] The quote is from Battigalli and Siniscalchi (2002, p. 374). For other relevant references, see e.g. Pearce (1984), Reny (1992), Asheim (2002), Perea (2014), Aumann and Arieli (2015), and Battigalli and De Vito (2018).

[4] 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 and Dufwenberg (2003) to whom we refer for more commentary and a model which shows how arguably attractive, epistemic assumptions ("common certain belief of full admissible consistency") admit play to leave the BI path.

[5] Up to here, the point was (essentially) made already by Luce and Raiffa (1957, pp. 80–81).

[6] See e.g. McKelvey and Palfrey (1992), Fey et al. (1996), Rapoport et al. (2003), Bornstein et al. (2004), Levitt et al. (2011), and the review Krockow et al. (2016). See also Binmore et al. (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).

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.[7] This allows us to explore the robustness of BI to concerns that some players may make a mistake.

We are furthermore 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).[8] We explore treatments reflecting that idea.

We do not attempt to exhaust the ways subjects may accumulate experience, however. For example, we do not have treatments allowing them to play the same game more than once random-matching style, an otherwise common experimenters' technique.[9] So, the "best shot" referred to in our paper's title refers to theoretical game-properties rather than lab-context.

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 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.

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. Finally, in each round, players are told whether they became the new king, but not what choice any particular co-player just made.

---

[7] 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 FRPD. See Petitt and Sugden (1989) for an early contribution, and Asheim and Dufwenberg (2003, section 4.3) for more.

[8] Dufwenberg et al. (2010) explore a similar issue in an otherwise different game (where the issue concerns the epiphany that one may have a dominant strategy). More broadly related are several other studies that one way or another study transfer of learning between games. See e.g. Nagel and Tang (1998), Cason et al. (2012), Cooper and Kagel (2008), Grimm and Mengel (2012), Haruvy and Stahl (2012), and Mengel and Sciubba (2014).

[9] A particular way of doing this was suggested by a referee: "I wonder whether it might be interesting to consider a treatment that uses Selten's [1967] 'strategy-method' [: ...] subjects first play the game a few times using the direct method to become familiar with the strategic environment, and only then submit a complete plan of action. This has the benefit that subjects get to submit a complete strategy (critical for BI) yet have some chance to become familiar with the strategic environment.".
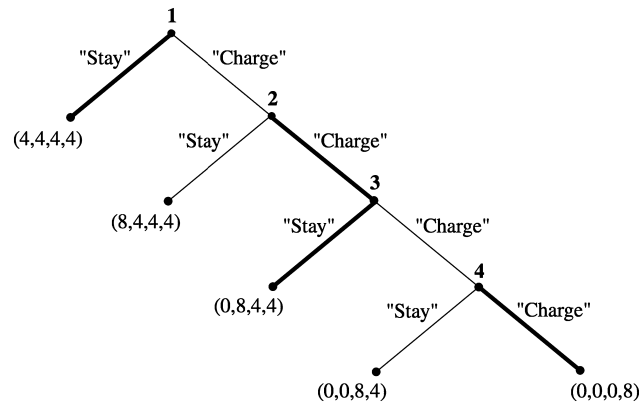
Fig. 1. Prediction line game ($N = 4$).

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. Adopting the notation C for "Charge the Hill" and S for "Stay Idle," 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. Fig. 1 illustrates the BI solution for the line game with four players.

Note that the prediction avoids the Basu–Reny critique,[10] described in the introduction. Following Battigalli & Siniscalchi one may also note that "rationality and $N - 1$ orders of mutual strong belief in mutual strong belief ... in rationality" implies the BI solution in a N-player KOH game. This suggests a sense in which the lower is N the stronger is the theoretical support for BI.

Robustness

BI may fail for many reasons. In particular, a player may be unable to backward induct, or he may deviate from BI not because he doesn't understand its logic but rather because he believes that others don't conform. A comparison between our line and ring games can shed light on that issue. We now present arguments that show that if subjects doubt whether others play BI, then it is harder to sustain BI in the ring game than in the line game.

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 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 $p \in [\frac{1}{2}, 1]$,[11] where p is independent and identical in each round.[12]

---

[10] This is obvious for all line games, since each player moves but once. For the ring games it is less clear, as players may move repeatedly. However, players never learn for sure that a co-player yet to move deviated from the BI path. 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).

[11] The assumption that $p \geq \frac{1}{2}$ is not necessary, but it rules out anomalous cases where players are not as likely to play BI. Since we are interested in the robustness of the BI it is natural to consider p in a neighborhood around $p = 1$.

[12] Heifetz and Pauzner (2005) explore a related model where with a low probability players gets "confused.".

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 $8(1 - p)^k$. 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

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

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 $8p^k$. 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

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

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

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

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 BI 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	$[0.7742, 1]$	$\left[\frac{1}{2}, 1\right]$
4	$[0.7742, 1]$	$\left[\frac{1}{2}, 1\right]$
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.

We finally establish that the limiting set of BI consistent beliefs for the ring game is the singleton set {1}.

Observation 3. *In the ring game, the set of BI consistent beliefs converges to {1} as $N \to \infty$.*

We can prove this observation is by contradiction. Suppose there exists $p < 1$ such that all KOH-ring games are BI consistent for all N for that p. Let "BI" be the odd-even strategy postulated to be optimal and consider some round $N-k$, where k is even. In this round, "BI" prescribes the action "Charge". Assume that the associated payoff from "BI" in the next round, if someone else became king, is v. Note that v must be at least 4. We show that "BI" cannot be optimal by comparing its payoff to that of the strategy S which chooses "Stay" at the current history but otherwise makes the same choices as "BI". Of course, a necessary condition for "BI" to be optimal is that it beats S.

Now, compare payoffs if we use "BI" vs. if we use S. The comparison can be structured over three exhaustive 'cases': First, suppose no other player charges in the current round. In this case "BI" brings $8p^k$ while S brings 4. So S does best if k is large enough. Second, suppose at least one other player charges in the current round but we become king if we charge. In this case "BI" brings $8p^k$ while S brings v. So, again, S does best if k is large enough. Third, suppose at least one other player charges in the current round but we do not become king if we charge. In this case "BI" brings v while S also brings v, a tie.

Each of these three cases occurs with positive probability (which can be calculated). Moreover, since we may choose k to be as large as we like by picking N to also be sufficiently large, we have shown that for large enough N and k, S beats "BI" in expectation. This is a contradiction.

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 also 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.[13]

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.

In summary, the experiment consisted of the following eight different treatments. In each cell we indicate the number of subjects that participated.

N	Ring game	Line game
3	23	21
4	19	21
2-then-3	24	17
2-then-4	18	20

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 a number corresponding 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.[14]

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: We asked subjects to read a short summary statement for each round and then to decide what they would do if placed in that situation, thus implementing a version of the strategy method.[15] This process allows us to observe behavior for information sets that may not be on the path of play. However, for each game, the process of eliciting contingent actions results in a different object. In the ring game the contingent decisions collected provide us with one strategy for each subject. In the line game, however, subjects make a contingent decision for each place in the line. Hence, the collection of actions provides a strategy for N different classes of line games – all the line games where the subject is player 1, all the line games where the subject is player 2, etc. Once all contingent decisions had been made, the experimenter collected the decision sheets from the subjects and randomly

---

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

[14] The instructions are available in the Appendix.

[15] We use the strategy method so that we may observe all choices on and off the path of play. However, we acknowledge that the salience of information may be different in a "live" version of the game.

Treatment	Strategy Choices																	BI	Other	Total

N=2

Treatment	ss	sc*	cs	cc														BI	Other	Total
Line 2	2	33	1	2														33	5	38
Ring 2	1	33	2	6														33	9	42

N=3

Treatment	sss	ssc	scs	css	ccs	csc*	scc	ccc										BI	Other	Total
Line 3	0	9	1	0	0	5	1	5										5	16	21
Line 3 (E)	0	7	0	0	0	10	0	0										10	7	17
Ring 3	2	16	0	0	0	1	2	2										1	22	23
Ring 3 (E)	0	12	0	1	0	10	1	0										10	14	24

N=4

Treatment	ssss	sssc	sscs	scss	csss	sscc	scsc*	sccs	cscs	cssc	ccss	sccc	cscc	ccsc	cccs	cccc	BI	Other	Total
Line 4	0	7	0	2	2	1	4	0	0	0	2	0	0	2	0	1	4	17	21
Line 4 (E)	0	8	0	0	0	0	9	0	0	0	0	1	0	0	0	3	9	12	21
Ring 4	5	9	1	0	0	2	0	0	0	0	0	0	0	1	0	1	0	19	19
Ring 4 (E)	0	11	0	0	0	1	3	0	0	0	0	0	1	1	0	1	3	15	18

*BI

Fig. 2. Individual strategy choice frequency.

matched people into groups and, in the case of the line game, randomly assigned player roles.[16] The game was then played out using the group's decisions.[17]

Subjects were paid at the end of the experiment as follows: For each subject, a group was constructed by combining the subject's own decision sheet with the decision sheets of the subjects with immediately higher number assignments (mod N). For example, if it was subject 3's turn to be paid for a four-player treatment, then his decision sheet was combined with the decision sheets of subjects 4, 5, and 6. This procedure was explained to subjects at the beginning of the experiment and allowed us to avoid the practical difficulty of forming groups if the total number of subjects in a session was not divisible by the correct number.[18] Subjects were privately paid their total earnings at the end of the session. In the line game treatments, subjects were only paid for one randomly selected game. Sessions typically lasted about a half hour. The average payout to subjects was about $10. No exchange rate was used.

3.3. Results

Fig. 2 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:

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, 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 $N \geq 3$.

---

[16] In the line game, a Bingo cage was used to assign subjects to player roles.

[17] In the ring game, if multiple subjects chose to charge the hill, then the king was selected at random using a Bingo cage.

[18] In treatments that involved two games the experimenter ensured that no player was matched with the same people twice.

Fifth, we look at the non-BI strategy choices made by subjects. Why did these subjects not play the BI prediction? We address two potential explanations. The first is that the decision to charge in a given round is risky in both versions of the KOH game. A subject who is risk averse may prefer to maximize their minimum payoff rather than play the BI prediction. We find a significant number of players who chose to play a maxmin strategy in the experiment. A second hypothesis that is suggested by the data is related to limited foresight.

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	Proportion BI			BI	Other	Proportion BI
Ring 2	33	9	0.79		Line 2	33	5	0.87
Ring 3	1	22	0.04		Line 3	5	16	0.24
Ring 4	0	19	0.00		Line 4	4	17	0.19
Ring 3 (E)	10	14	0.42		Line 3 (E)	10	7	0.59
Ring 4 (E)	3	15	0.17		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).

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 previous 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.[19] 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	Comparison	p-value
Ring 2 vs. Ring 3	0.000**	Ring 2 vs. Ring 4	0.000**
Ring 3 vs. Ring 4	0.548	Ring 3 (E) vs. Ring 4 (E)	0.080*
Line 2 vs. Line 3	0.000**	Line 2 vs. Line 4	0.000**
Line 3 vs. Line 4	0.500	Line 3 (E) vs. Line 4 (E)	0.257

There is some evidence to support the hypothesis that games with smaller *N* will have a higher proportion of players who played according to the theoretical prediction. We strongly reject equality of proportions of the two- and three-player treatments for both games in favor of the one-sided alternative and equality of proportions of the two- and four-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.

---

[19] See, for instance, Siegal and Castellan (1988, p. 103). While we use the Fisher test in our paper, this test makes certain questionable assumptions about the data. The Fisher test is used when two independent samples can fall into one of two mutually exclusive classes. In order to compute the test statistic, it holds the population of the two samples as fixed (which is fixed by us) as well as the total number of observed "successes." This later marginal is random and not controlled by the experimenter. In contrast, the Barnard (1945) test is also an exact test for the same environment, but it only conditions on the population of the two samples. That said, while the Barnard test avoids conditioning on the total number of observations in each category, it is computationally intensive to generate the test statistic as well as difficult to determine the appropriate tail events to implement a two sided test. We have therefore opted to use the Fisher's exact test. We are grateful to an anonymous referee for bringing the Barnard test and the issues with Fisher's exact test to our attention.

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 statistically significant at the 5% level. In the four-player treatments, the difference is not significant at that level. In both of the experienced treatments, players were first exposed to a two-player game. 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.

Results, Part 5: Where and How does BI breakdown?

In the previous results we have compared the proportion of subjects who played the BI prediction across group sizes and games. BI does well in the two-player games. However, there is a quick breakdown in the prediction for both games as the number of players goes from two to three even in the treatments where players are given experience. What is going on? While the purpose of the experiment was to provide a clean test of the BI hypothesis, in this section we use the experimental data for each game in order to shed some potential insight into "where" and "how" the BI prediction breaks in our experiment.

Fig. 2 contains the complete set of contingent decisions made by all of the subjects. We briefly comment on one of the "other" strategy choices subjects made in the experiment. In particular, the most commonly played strategy in almost all of the treatments was one where the subject chose S in all rounds except the last round where they play C. This is a player's maxmin strategy in both of the KOH games.[20] In fact, it is not only maxmin, but the restriction of this strategy to

---

[20] In the ring game this is the maxmin strategy. In the line game, this profile is the collection of maxmin strategies – the maxmin strategy for each player position.

Treatment	Strategy Choices				BI	Other	Total	%BI
N=3, Round 2 (Restriction)	ss	sc*	cs	cc				
Line 3	0	14	1	6	14	7	21	0.67
Line 3 (E)	0	17	0	0	17	0	17	1.00
Ring 3	2	17	0	4	17	6	23	0.74
Ring 3 (E)	1	22	0	1	22	2	24	0.92
N=3, Round 3 (Restriction)	s	c*						
Line 3	1	20			20	1	21	0.95
Line 3 (E)	0	17			17	0	17	1.00
Ring 3	2	21			21	2	23	0.91
Ring 3 (E)	1	23			23	1	24	0.96

Fig. 3. Individual strategy choice frequency for $N = 3$ treatments (restrictions).

each subgame is also maxmin for the subgame. Hence, the strategy is maxmin perfect. It maximizes the player's minimum possible payoff at each point in the game. In short, maxmin is a conservative, but rational, way to play the game if a subject is risk averse. The table below compares the proportion of the observations consistent with maxmin play vs. BI play in each treatment.

	Proportion				Proportion	
	Maxmin	BI			Maxmin	BI
Ring 2	0.79	0.79		Line 2	0.87	0.87
Ring 3	0.70	0.04		Line 3	0.43	0.24
Ring 4	0.47	0.00		Line 4	0.33	0.19
Ring 3 (E)	0.50	0.42		Line 3 (E)	0.41	0.59
Ring 4 (E)	0.61	0.17		Line 4 (E)	0.38	0.43

A couple of observations are apparent from this table. First, the combination of BI and maxmin organize a significant proportion of the observed data in most cases. In the two-player games, maxmin is the same as BI hence the observed proportions are the same. However, for all other game sizes, the two solution concepts differ in their prediction. Second, the frequency of maxmin play relative to the BI prediction is quite large in the ring games. In fact, the proportion of maxmin play is higher in the ring game than the line game for all group sizes larger than two. This is logical since, as we have argued earlier, charging in the ring game is riskier than charging in the line game. Hence, the choice of maxmin in the ring game makes more sense if the subject is risk averse.

Finally, we look at the restrictions of the subjects' strategies to each round. The motivation for looking at the strategies in this way is as follows: some subjects might not backward induct immediately. It may only occur to them later in the game that they should look forward and reason back. If this is the case, then what we should observe is that the restriction of their strategy to later rounds should coincide with the restriction of the BI prediction to later rounds. These adjustments are missed by only looking the strategy for the whole game. Figs. 3 and 4 break down the empirical choices from the three and four-player games respectively into their restrictions to each round.

Thus, for each round, we may ask whether a subject's continuation strategy matches the BI prediction for the remainder of the game. The table below presents the proportion of the restricted strategies that coincide with the BI prediction for the three-player games.

Proportion BI for Three-player Games (by Round Restriction)			
	Round 1	Round 2	Round 3
Line 3	0.24	0.67	0.95
Line 3 (E)	0.59	1.00	1.00
Ring 3	0.04	0.74	0.91
Ring 3 (E)	0.42	0.92	0.96

The analogous table for the four-player games is provided below.

Treatment	Strategy Choices								BI	Other	Total	%BI
N=4, Round 2 (Restriction)	sss	ssc	scs	css	ccs	csc•	scc	ccc				
Line 4	2	7	0	4	0	6	1	1	6	15	21	0.29
Line 4 (E)	0	8	0	0	0	9	0	4	9	12	21	0.43
Ring 4	5	9	1	0	0	1	2	1	1	18	19	0.05
Ring 4 (E)	0	11	0	0	0	4	2	1	4	14	18	0.22
Round 3 (Restriction)	ss	sc•	cs	cc								
Line 4	6	13	0	2					13	8	21	0.62
Line 4 (E)	0	17	0	4					17	4	21	0.81
Ring 4	5	10	1	3					10	9	19	0.53
Ring 4 (E)	0	15	0	3					15	3	18	0.83
Round 4 (Restriction)	s	c•										
Line 4	6	15							15	6	21	0.71
Line 4 (E)	0	21							21	0	21	1.00
Ring 4	6	13							13	6	19	0.68
Ring 4 (E)	0	18							18	0	18	1.00

Fig. 4. Individual strategy choice frequency for $N = 4$ treatments (restrictions).

Proportion BI for Four-Player Games
(by Round Restriction)

	Round 1	Round 2	Round 3	Round 4
Line 4	0.19	0.29	0.62	0.71
Line 4 (E)	0.43	0.43	0.81	1.00
Ring 4	0.00	0.04	0.53	0.68
Ring 4 (E)	0.17	0.22	0.83	1.00

In each treatment, we observe the proportion of BI play in both the three- and four-player games is increasing with each subsequent restriction. The increase in % BI at each restriction means that at least one subject, who did not play the BI predicted action in the previous round, has switched to the BI prediction for the remainder of the game. This observation is in line with our previous comparisons of the separate treatments involving different group sizes.

Finally, while our primary motivation for developing the model of BI consistency was to be able to make predictions between the ring and line games, the model generates several other interesting predictions. In the ring game, for instance, the model suggests that the range of p that are BI consistent is decreasing in N. In particular, a subject may have a belief p that makes BI optimal in later rounds and suboptimal in earlier rounds. Hence, if p are equally likely, subjects' in the ring game are more likely to play the BI prediction the closer they are to the end of the game. This prediction is supported by the restriction data. However, in the line game, the range of BI consistent beliefs is constant. Thus, if a subject has a belief p that makes it optimal for him to play the BI prediction in the last round, then he should play the BI prediction in all previous rounds. In other words, we do not expect the proportion of BI play in the line game to be increasing with each restriction. The data therefore does not seem to support the iid model's prediction for the line game. Consequently, the strategic uncertainty built into our iid model is not enough to explain the observation that the proportion of subjects playing the BI prediction is decreasing with the number of players in both games. This prediction for the line game is a consequence of our strong assumption that each player i believes that the other players choose the BI predicted action independently with the same probability p in all rounds (i.e., p does not vary with the number of players left in the game). We used this assumption to make our model more tractable and give us sharper predictions. In a more realistic model, one could allow for p to vary with the number of players remaining in the game (e.g., p decreasing). The added degrees of freedom in the model would accommodate a wider range of predictions, but at the cost of decreased tractability.

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 not 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 et al. (2013) for a survey.[21] 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, 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 et al. (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 et al. (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.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.geb.2018.07.007.

References

Asheim, G., 2002. On the epistemic foundation for backward induction. Math. Soc. Sci. 44, 121–144.
Asheim, G., Dufwenberg, M., 2003. Deductive reasoning in extensive form games. Econ. J. 113, 305–325.
Bar-Gill, O., Persico, N., 2016. Exchange efficiency with weak ownership rights. Am. Econ. J. Microecon. 8, 230–267.
Barnard, G.A., 1945. A new test for 2 × 2 tables. Nature 156, 177.

---

[21] Important contributions getting this literature started include Stahl and Wilson (1994, 1995), Nagel (1995), Costa-Gomes et al. (2001), and Camerer et al. (2004).

Basu, K., 1988. Strategic irrationality in extensive games. Math. Soc. Sci. 15, 247–260.

Battigalli, P., De Vito, N., 2018. Beliefs, Plans, and Perceived Intentions in Dynamic Games. Bocconi University. Unpublished manuscript.

Battigalli, P., Siniscalchi, M., 2002. Strong belief and forward induction reasoning. J. Econ. Theory 16, 356–391.

Ben-Porath, E., 1997. Rationality, Nash equilibrium, and backwards induction in perfect information games. Rev. Econ. Stud. 64, 23–46.

Bernheim, D., 1984. Rationalizable strategic behavior. Econometrica 52, 1007–1028.

Binmore, K., 1987. Modeling rational players: Part I. Econ. Philos. 3, 179–214.

Binmore, K., McCarthy, J., Ponti, G., Samuelson, L., Shaked, A., 2001. A backward induction experiment. J. Econ. Theory 104, 48–88.

Bornstein, G., Kugler, T., Ziegelmeyer, A., 2004. Individual and group decisions in the centipede game: are groups more 'rational' players?. J. Exp. Soc. Psychol. 40, 599–605.

Brams, S., Kilgour, D., 1998. Backward induction is not robust: the parity problem and the uncertainty problem. Theory Decis. 45, 263–289.

Camerer, C., 2003. Behavioral Game Theory. Russell Sage Foundation.

Camerer, C., Ho, T.-H., Chong, J.-K., 2004. A cognitive hierarchy model of games. Q. J. Econ. 119, 861–898.

Cason, T., Savikhon, A., Sheremata, R., 2012. Behavioral spillovers in coordination games. Eur. Econ. Rev. 56, 233–245.

Cooper, D., Kagel, J., 2008. Learning and transfer in signaling games. Econ. Theory 34, 415–439.

Costa-Gomes, M., Crawford, V., Broseta, B., 2001. Cognition and behavior in normal-form games: an experimental study. Econometrica 69, 1193–1235.

Costa-Gomes, M., Crawford, V., Irriberri, N., 2013. Structural models of nonequilibrium strategic thinking: theory, evidence, and applications. J. Econ. Lit. 51, 5–62.

Dufwenberg, M., Sundaram, R., Butler, D., 2010. Epiphany in the game of 21. J. Econ. Behav. Organ. 75, 132–143.

Fey, M., McKelvey, R., Palfrey, T., 1996. An experimental study of constant-sum centipede games. Int. J. Game Theory 25, 269–287.

Friedenberg, A., Kets, W., Kneeland, T., 2016. Bounded Reasoning: Rationality or Cognition. ASU, Northwestern, and UCL. Unpublished manuscript.

Grimm, V., Mengel, F., 2012. An experiment on learning in multiple games environment. J. Econ. Theory 147, 2220–2259.

Gul, F., 1997. Rationality and coherent theories of strategic behavior. J. Econ. Theory 70, 1–31.

Johnson, E., Camerer, C., Sen, S., Rymon, T., 2002. Detecting failures of backward induction: monitoring information search in sequential bargaining. J. Econ. Theory 104, 16–47.

Haruvy, E., Stahl, D., 2012. Between-game rule learning in dissimilar symmetric normal-form games. Games Econ. Behav. 74, 208–221.

Heifetz, A., Pauzner, A., 2005. Backward induction with players who doubt others' faultlessness. Math. Soc. Sci. 50, 252–267.

Kaplow, L., Shavell, S., 1996. Property rules versus liability rules. An economic analysis. Harvard Law Rev. 109, 713–790.

Krockow, E., Colman, A., Pulford, P., 2016. Cooperation in repeated interactions: a systematic review of centipede game experiments, 1992–2016. Eur. Rev. Soc. Psychol. 27, 231–282.

Kneeland, T., 2015. Identifying higher-order rationality. Econometrica 83, 2065–2079.

Levitt, S., List, J., Sadoff, S., 2011. Checkmate: exploring backward induction among chess players. Am. Econ. Rev. 101, 975–990.

Luce, D., Raiffa, H., 1957. Games and Decisions. Wiley, New York.

Mengel, F., Sciubba, E., 2014. Extrapolation and structural similarity in games. Econ. Lett. 125, 381–385.

Mantovani, M., 2014. Limited Backward Induction: Foresight and Behavior in Sequential Games. University of Milan Bicocca. Unpublished manuscript.

McKelvey, R., Palfrey, T., 1992. An experimental study of the centipede game. Econometrica 60, 803–836.

Nagel, R., 1995. Unraveling in guessing games: an experimental study. Am. Econ. Rev. 85, 1313–1326.

Nagel, R., Tang, F., 1998. Experimental results on the centipede game in normal form: an investigation on learning. J. Math. Psychol. 42, 356–384.

O'Donoghue, T., Rabin, M., 1999. Doing it now or later. Am. Econ. Rev. 89, 103–124.

Pearce, D., 1984. Rationalizable strategic behavior and the problem of perfection. Econometrica 52, 1029–1050.

Perea, A., 2014. Belief in the opponents' future rationality. Games Econ. Behav. 83, 231–254.

Petitt, P., Sugden, R., 1989. The backward induction paradox. J. Philos. 4, 169–182.

Rapoport, A., Stein, W., Parco, J., Nicholas, T., 2003. Equilibrium play and adaptive learning in a three-person centipede game. Games Econ. Behav. 43, 239–265.

Reny, P., 1988. Rationality, Common Knowledge and the Theory of Games. PhD Dissertation. Chapter 1. Department of Economics, Princeton University.

Reny, P., 1992. Backward induction, normal form perfection and explicable equilibria. Econometrica 60, 627–649.

Reny, P., 1993. Common belief and the theory of games with perfect information. J. Econ. Theory 59, 257–274.

Roomets, A., 2010. On Limited Foresight in Games. University of Arizona. Unpublished manuscript.

Rosenthal, R., 1981. Games of perfect information, predatory pricing and the chain-store paradox. J. Econ. Theory 25, 92–100.

Selten, R., 1967. Die Strategiemethode zur Erforschung des Eingeschrankt Rationalen Verhaltens im Rahmen eines Oligopolexperiments. In: Sauermann, H. (Ed.), Beitrage zur Experimentellen Wirtschaftsforschung. J.C.B. Mohr, Tubingen, pp. 136–168.

Selten, R., 1978. The chain store paradox. Theory Decis. 9, 127–159.

Siegel, C., Castellan, J., 1988. Nonparametric Statistics for the Behavioral Sciences, 2nd ed.. McGraw-Hill.

Stahl, D., Wilson, P., 1994. Experimental evidence on player's models of other players. J. Econ. Behav. Organ. 25, 309–327.

Stahl, D., Wilson, P., 1995. On player's models of other players: theory and experimental evidence. Games Econ. Behav. 10, 218–254.

Stewart, I., 1999. A puzzle for pirates. Sci. Am. 280 (5), 98–99.