2-2013

A Dynamic Level-k Model in Sequential Games

Teck-Hua Ho

Xuanming Su
University of Pennsylvania

Recommended Citation
Ho, T., & Su, X. (2013). A Dynamic Level-k Model in Sequential Games. *Management Science, 59* (2), 452-469. http://dx.doi.org/10.1287/mnsc.1120.1645

A Dynamic Level-k Model in Sequential Games

Abstract

Backward induction is a widely accepted principle for predicting behavior in sequential games. In the classic example of the "centipede game," however, players frequently violate this principle. An alternative is a "dynamic level-k" model, where players choose a rule from a rule hierarchy. The rule hierarchy is iteratively defined such that the level-k rule is a best response to the level-(k-1) rule, and the level-∞ rule corresponds to backward induction. Players choose rules based on their best guesses of others' rules and use historical plays to improve their guesses. The model captures two systematic violations of backward induction in centipede games, limited induction and repetition unraveling. Because the dynamic level-k model always converges to backward induction over repetition, the former can be considered to be a tracing procedure for the latter. We also examine the generalizability of the dynamic level-k model by applying it to explain systematic violations of backward induction in sequential bargaining games. We show that the same model is capable of capturing these violations in two separate bargaining experiments.

Keywords

level-k models, learning, sequential games, backward induction, behavioral game theory

Disciplines

Finance and Financial Management | Social and Behavioral Sciences

A Dynamic Level-k Model in Games

Teck-Hua Ho and Xuanming Su*

September 21, 2010

Backward induction is the most widely accepted principle for predicting behavior in dynamic games. In experiments, however, players frequently violate this principle. An alternative is a 2-parameter "dynamic level-k" model, where players choose a rule from a rule hierarchy. The rule hierarchy is iteratively defined such that the level-k rule is a best-response to the level-$(k-1)$ rule and the level-∞ rule corresponds to backward induction. Players choose rules based on their best guesses of others' rules and use past plays to improve their guesses. The model captures two systematic violations of backward induction, namely limited induction and time unraveling, and helps to resolve paradoxical behaviors in the centipede game, finitely repeated prisoner's dilemma, and chain store game, three canonical games where backward induction performs poorly. The dynamic level-k model can be considered as a tracing procedure for backward induction because the former converges to the latter in the limit.

Keywords: Level-k Models, Learning, Dynamic Games, Backward Induction, Behavioral Game Theory, Experimental Economics

*Haas School of Business, University of California at Berkeley. Authors are listed in alphabetical order. We thank Colin Camerer and Juin-Kuan Chong for collaboration in the project's early stages, and Vince Crawford for extremely helpful comments. We are also grateful for constructive feedback from seminar participants at University of California at Berkeley, University of Pennsylvania, and University of Oxford. Direct correspondence to any of the authors. Ho and Su: Haas School of Business, University of California at Berkeley, Berkeley, CA 94720-1900, Email: Ho: hoteck@haas.berkeley.edu; Su: xuanming@haas.berkeley.edu.

I. Introduction

Players often interact with one another over multiple stages in many economic situations. Economists invoke the principle of backward induction to predict behavior in these games. Under backward induction, players reason backward, replace each subgame by its optimal payoff, always choose rationally at each subgame, and use this iterative process to determine a sequence of optimal actions. Each player follows this procedure betting on others doing the same. This "divide and conquer" algorithm simplifies the game analysis and generates a sharp prediction of game play at every subgame.

However, there are several canonical games showing that people often violate backward induction. One of these games is the so-called "centipede" game (Rosenthal, 1981) (See top panel of Figure 1). In this game, there are 2 players (A and B) and 4 decision stages. Players are endowed with an initial pot of \$5. In Stage I, Player A has the property rights to the pot. She can choose either to end the game by taking 80% of the pot (and leaving the remaining 20% to Player B) or to allow the pot to double by passing the property rights to Player B. In Stage II, it is now Player B's turn to make a similar decision. Player B must now decide whether to end the game by taking 80% of \$10 or to let the pot double again by passing the property rights back to Player A. This social exchange process leads to large financial gains as long as both players surrender their property rights at each stage. At Stage IV, Player B can either take 80% of \$40 (i.e., \$32) or pass and be left with 20% of a pot of \$80 (i.e., \$16).

The principle of backward induction generates a sharp prediction for this game by starting the analysis in the very last stage. Player B should take at Stage IV because 80% of \$40 is larger than 20% of \$80. Anticipating this choice, Player A should take at Stage III since 80% of \$20 is larger than 20% of \$40. Continuing with this line of logic, backward induction makes a surprising prediction: Player A always takes immediately in Stage I and Outcome 4 occurs with probability 1 (i.e. Outcomes 0-3 should not occur). Moreover, the same prediction holds even if the game continues for more stages and with more dramatic potential gains. For example, the bottom panel of Figure 1 shows the same game with 6 stages and the same sharp prediction holds (i.e., Outcome 6 occurs with probability 1). Introspection suggests that this prediction is unlikely to occur. This is so because as long as the game proceeds to Stage III, both players would have earned more money than the backward induction outcome.

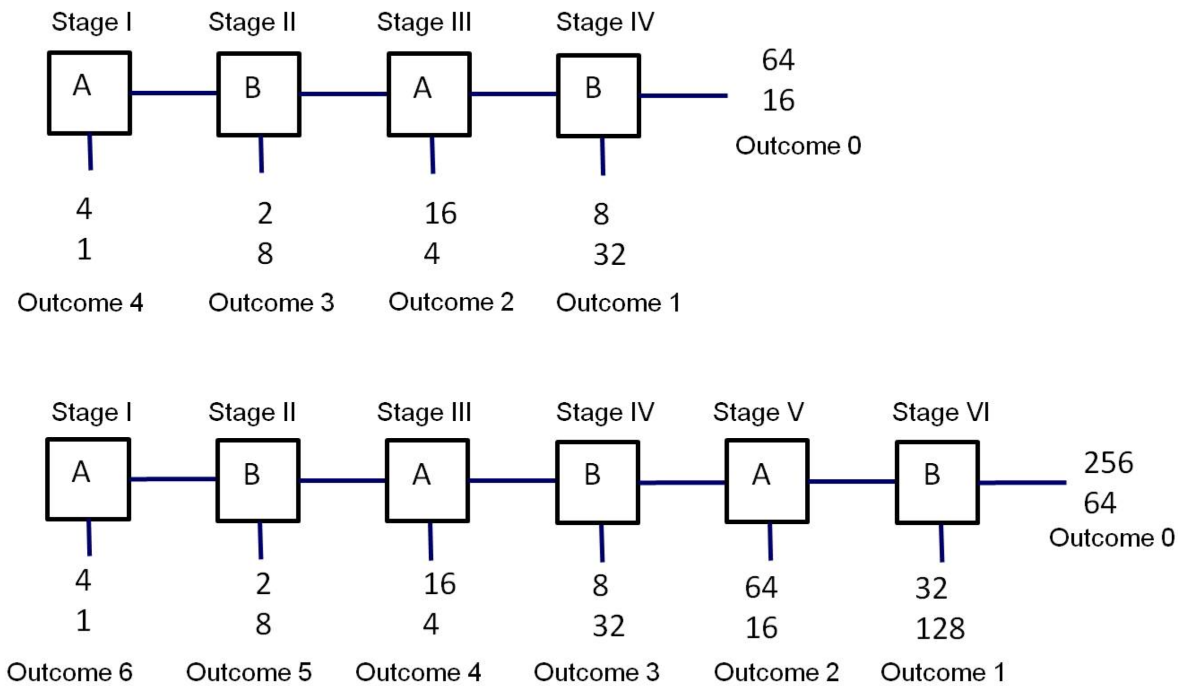

Figure 1: 4-stage (top) and 6-stage (bottom) centipede games.

Indeed, very few subjects (about 6%) choose to take immediately in experimental centipede games conducted by McKelvey and Palfrey (1992). Many subjects choose to take in the intermediate stages, approximately half way through the games (i.e., leading to Outcomes 2-3 in 4-stage and Outcomes 3-4 in 6-stage games). Clearly, this pattern of behavior runs counter to backward induction. Moreover, the observed behavior frequently leads to higher cash earnings for all subjects. Other well-known games where backward induction performs poorly include the repeated prisoner's dilemma (Axelrod, 1984) and the chain store market entry game (Selten, 1978). In the latter two cases, the unique prediction prescribed by backward induction was rarely played especially when subjects were inexperienced.

There are two stylized facts concerning the violations of backward induction. First, players violate backward induction less in a game with fewer subgames (or stages). That is, players' behaviors deviate less from backward induction in simpler games. For instance, we observe fewer violations of backward induction in 4-stage than in 6-stage game (see Figure 1). We call this behavioral tendency *limited induction.* Second, players unravel over time as they play the same game repeatedly over multiple rounds. That is, players' behaviors converge towards backward induction over time. For instance, we observe fewer violations of backward induction in the last round than in the first round of the experimental centipede games in Figure 1. This behavioral tendency is termed *time unraveling*. The inability of backward induction to account for the two empirical stylized facts poses modeling challenges for economists.

In this paper, we propose an alternative to backward induction, a "dynamic level-k" model, that generalizes backward induction and accounts for limited induction and time unraveling. In the dynamic level-k model, players choose a level-k rule, L_k $(k = 0, 1, 2, 3 \ldots)$, from a set of iteratively defined rules and the chosen rule prescribes an action at each subgame (Stahl and Wilson, 1995; Stahl, 1996; Ho et al. 1998; Costa-Gomes et al. 2001; Costa-Gomes and Crawford, 2006; Crawford and Iriberri, 2007a, 2007b). The rule hierarchy is defined such that the level-k rule best-responds to the level-$(k-1)$ rule and the level-∞ rule corresponds to backward induction. Since players choose a rule based on their beliefs of others' rules, they essentially are subjective expected utility maximizers.

Players are heterogenous in that they have different initial guesses of others' rules and consequently choose different initial rules. The distribution of the initial guesses is assumed to follow a Poisson distribution. These initial guesses are updated according to Bayes' rule based on game history (cf. Stahl, 2000). Consequently, players develop more accurate guess of others' rules and may choose different rules over time.

We prove that the dynamic level-k model can account for limited induction and time unraveling properties in the centipede game, the finitely repeated prisoner's dilemma, and the chain store game. Consequently our model can explain passing in the centipede game, cooperation in the repeated prisoner's dilemma, and fighting by the incumbent in the chain-store game. All these behaviors are considered paradoxical under backward induction but are predicted by the dynamic level-k model. In addition, the dynamic level-k model is able to capture the empirical stylized fact that behavior will eventually converge to backward induction over time, and hence the former can be considered as a tracing procedure for the latter.

We fit our model using experimental data on the centipede game from McKelvey and Palfrey

(1992) and find that our model fits the data significantly better than backward induction and static level-k model. We rule out two alternative explanations including the reputation-based model of Kreps et al. (1982) and a model allowing for social preferences (Fehr and Schmidt, 1999). Overall, it appears that the dynamic level-k model can be an empirical alternative to backward induction.

The rest of the paper is organized as follows. Section II discusses the backward induction principle and its violations. Section III formulates the dynamic level-k model and applies it to explain paradoxical behaviors in the centipede game, iterated prisoner's dilemma, and chain store game. Section IV fits the dynamic level-k model to data from experimental centipede game and rules out two alternative explanations. Section V concludes.

II. Violations of Backward Induction

Backward induction uses an iterative process to determine an optimal action at each subgame. The predictive success of this iterative reasoning process hinges on players' complete confidence in others applying the same logic in arriving at the backward induction outcomes (Aumann, 1995). If players have doubts about others applying backward induction, it may be in their best interest to deviate from the prescription of backward induction. Indeed, subjects do and profitably so in many experiments.

If a player i chooses a behavioral rule L^i that is different from backward induction (L_∞), one would like to develop a formal measure to quantify this deviation. Consider a game G with S subgames. We can define the deviation for a set of behavioral rules $L^i (i = 1, \ldots I)$, one for each player, in game G as:

$$\delta(L^1, \ldots, L^I, G) = \frac{1}{S} \sum_{s=1}^{S} \left[\frac{1}{N_s} \sum_{i=1}^{N_s} D_s(L^i, L_\infty) \right], \tag{II.1}$$

where $D_s(L^i, L_\infty)$ is 1 if player i chooses an action at subgame s that is different from the prescription of backward induction and 0 otherwise, and N_s is the number of players who are active at subgame s. Note that the measure varies from 0 to 1, where 0 indicates that players' actions perfectly match the predictions of backward induction and 1 indicates that none of the players' actions agree with the predictions of backward induction.

Let us illustrate the deviation measure using a 4-stage centipede game. Let the behavioral rules adopted by player A and B be $L^A = \{P, -, T, -\}$ and $L^B = \{-, P, -, T\}$ respectively (that is, player A will pass in Stage I and take in Stage III, and player B will pass in Stage II and take in Stage IV). Then the game will end in Stage III (i.e., Outcome 2). The deviation will be $\delta(L^A, L^B, G) = \frac{1}{4}[1 + 1 + 0 + 0] = \frac{1}{2}$. Similarly, if $L^A = \{P, -, T, -\}$ and $L^B = \{-, T, -, T\}$, then the game will end in Stage II (i.e., Outcome 3). This gives $\delta(L^A, L^B, G) = \frac{1}{4}[1 + 0 + 0 + 0] = \frac{1}{4}$, which is smaller. Note that the latter behavioral rules are closer to backward induction than the former behavioral rules.

Using the above deviation measure, we can formally state the two systematic violations of backward induction as follows:

1. *Limited Induction*: Consider two games G and G' where G' is a proper subgame of G. The deviation from backward induction is equal or larger in G than in G'. That is, the deviation from backward induction increases in S. Formally, for a group of players who adopt the same set of behavioral rules ($L^i, i = 1, \ldots, I$) in games G and G', we have $\delta(L^1, \ldots, L^I, G) \geq \delta(L^1, \ldots, L^I, G')$. Consequently, a good model must predict a larger deviation in G than in G' to be behaviorally plausible.

2. *Time Unraveling*: If a game G is played repeatedly over time, the deviation from backward induction at time t converges to zero as $t \to \infty$. That is, time unraveling implies $\delta(L^1(t), \ldots, L^I(t), G) \to 0$ as $t \to \infty$. Therefore, game outcomes will eventually be consistent with backward induction after sufficiently many repetitions.

Let us illustrate limited induction and time unraveling using data from McKelvey and Palfrey (1992). These authors conducted an experiment to study behavior in 4-stage and 6-stage centipede games. Each subject was assigned to one of these games and played the same game in the same role 9 or 10 times. For each observed outcome in a game play, we can compute the deviation from backward induction using Equation (II.1).[1]

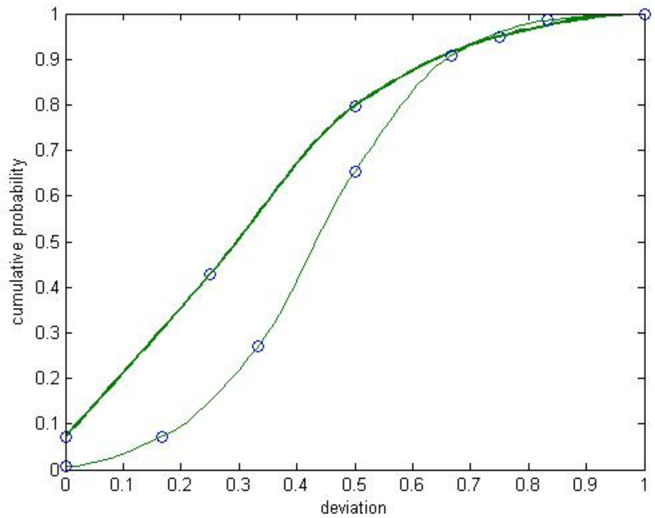

Figure 2: Deviations from backward induction in the 4-stage and 6-stage centipede games (Data from McKelvey and Palfrey, 1992).

Figure 2 plots the cumulative distributions of deviations from backward induction in the 4-stage and 6-stage games respectively. The thick line corresponds to the 4-stage game and the thin line corresponds to the 6-stage game. The curve for the 6-stage game generally lies to the right of the curve for the 4-stage game except for high deviation values. A simple t-test shows that there is a statistically significant difference between the mean deviations in the two games. These results suggest that the limited induction property holds in this data set.

Figure 3 plots the cumulative distributions of deviations from backward induction in the first and the final round of the 4-stage game. The thick line corresponds to the first round and the thin line corresponds to the final round of game plays (similar results occur for the

[1]Since subjects did not indicate what they would have chosen in every stage (i.e., data were not collected using the strategy method), we do not observe what the subjects would have done in subsequent stages if the game ended in an earlier stage. In computing the deviation, we assume that subjects always choose to take in stages beyond where the game ends. Therefore, the derived measure is a conservative estimate of the deviation from backward induction.

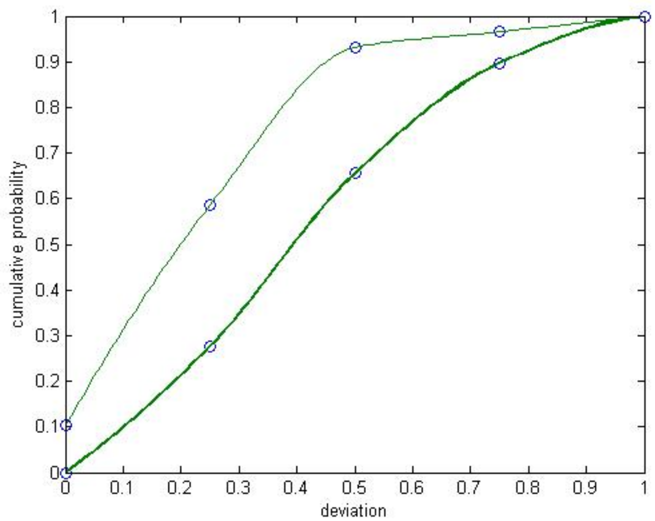

Figure 3: Deviations from backward induction in the first and the last round of the 4-stage centipede game (Data from McKelvey and Palfrey, 1992).

6-stage game). As shown, the curve for the first round lies to the right of the curve for the final round. A simple t-test shows that there is a statistically significant difference between the mean deviations in the first and last rounds. These results suggest that the deviation from backward induction decreases over time.

We shall use the deviation measure to establish the main theoretical results below. Specifically, we shall show that the deviation under dynamic level-k model is smaller in simpler games (i.e., limited induction property holds) and converges to zero over time (i.e., satisfies the time unraveling property).

III. Dynamic Level-k Model

We consider a 2-player game that has S subgames. Players are indexed by i ($i = 1, 2$) and subgames by s ($s = 1, \ldots, S$). Players are assumed to adopt a rule which prescribes an action at each subgame s. For example, in the centipede game studied by McKelvey and Palfrey (1992), S is either 4 or 6 and a rule Player A adopts in a 4-stage game can be $L^A = \{P, -, T, -\}$, which specifies that the player will pass in Stage I and take in Stage III.

Players choose a rule from a rule hierarchy. Rules are denoted by L_k ($k = 0, 1, 2, 3, \ldots$) and are generated from iterative best-responses. In general, L_k is a best-response to L_{k-1} at every subgame, and L_∞ corresponds to backward induction. Under this rule hierarchy, the deviation from backward induction (see equation (I.1)) is smaller if player i adopts a higher level rule (while others keep their rule at the same level). In fact, *ceteris paribus*, the deviation from backward induction is monotonically decreasing in k. In other words, the level of a player's rule captures its closeness to backward induction.

Note that L_k prescribes the same behavior as backward induction in any game with k or fewer subgames. In this regard, L_k can be viewed as a limited backward induction rule that only works for simpler games. Putting it this way, a higher level rule is more likely to coincide with backward induction in a wider class of games so that the level of a rule measures its degree of resemblance to backward induction.

If every player is certain about others' rationality, all players will choose L_∞. However, if players have doubts about others' rationality, it is not in their best interest to apply backward induction. Instead, they should form beliefs over rules adopted by others[2] and choose a best-response rule in order to maximize their expected payoffs. Therefore, under the model, players are subjective expected utility maximizers. In this regard, the dynamic level-k model is similar to the notion of rationalizable strategic behavior (Bernheim, 1984; Pearce, 1984; Reny, 1992) except that players always choose a rule from the defined rule hierarchy.

In a typical laboratory experiment, players frequently play the same game repeatedly. After each game round, players observe the rules used by their opponents and update their beliefs by tracking the frequencies of rules played by opponents in the past. Let player i's rule counts at the end of round t be $N^i(t) = (N^i_0(t), \ldots, N^i_S(t))$ where $N^i_k(t)$ is the cumulative count of rule L_k that has been used by opponents at the end of round t (Camerer and Ho, 1999; Ho et

[2]Subjects' beliefs may depend on their knowledge of their opponents' rationality. Players who play against opponents who are known to be sophisticated will adopt a higher level rule. For example, Palacios-Huerta and Volijc (forthcoming) show that many chess players in the centipede game choose to pass when playing against student subjects but they choose to take immediately when playing against other equally sophisticated chess players. Levitt et al. (2009) however find chess players choose to pass when they play against each other. This discrepancy could be due to the difference in players' perception of their opponents' rationality.

al., 2007). Note that for a game with S subgames, all rules of level S or higher will prescribe the same action at each subgame and hence we pool them together and collectively call them L_S. Given these rule counts, player i forms a belief $B^i(t) = (B^i_0(t), \ldots, B^i_S(t))$ where

$$B^i_k(t) = \frac{N^i_k(t)}{\sum_{k'=0}^{S} N^i_{k'}(t)}. \tag{III.1}$$

$B^i_k(t)$ is player i's belief of the probability that her opponent will play L_k in round $t+1$. The updating equation of the cumulative count at the end of round t is given by:

$$N^i_k(t) = N^i_k(t-1) + I(k,t) \cdot 1, \ \ \forall k \tag{III.2}$$

where $I(k,t) = 1$ if player i's opponent adopts rule L_k in round t and 0 otherwise. Therefore, players update their beliefs based on the history of game plays. This updating process is consistent with Bayesian updating involving a multinomial distribution with a Dirichlet prior (Fudenberg and Levine, 1998, Camerer and Ho, 1999). As a consequence of the updating process, players may adopt a different best-response rule in round $t+1$.[3]

[3]The above updating rule assumes that subjects observe rules chosen by opponents. This is possible if the strategy method is used to elicit subjects' contingent action at each subgame. When the opponents' chosen rules are not observed, the updating process is still a good approximation because subjects may have a good guess of their opponents' chosen rules in most simple games (e.g., centipede games). More generally, the updating of $N^i_k(t)$ depends on whether player i adopts a higher or lower level rule than her opponent. If the opponent uses a higher level rule (e.g., the opponent takes before the player in the centipede game), then we have similar to the above:

$$N^i_k(t) = N^i_k(t-1) + I(k,t) \cdot 1 \tag{III.3}$$

where $I(k,t) = 1$ if opponent adopts an action that is consistent with L_k in round t and 0 otherwise. If player i adopts a higher level rule k^* (e.g., takes before the opponent in the centipede game), the player can only infer that the opponent has chosen some rule that is below k^*. Then we have:

$$N^i_k(t) = N^i_k(t-1) + I(k \leq k^*) \cdot \frac{N^i_k(t-1)}{\sum_{k'=0}^{k^*} N^i_{k'}(t-1)} \tag{III.4}$$

where $I(k \leq k^*) = 1$ if $k \leq k^*$ and 0 otherwise. This updating process assigns a belief weight to all lower level rules that are consistent with the observed outcome. The weight assigned to each consistent rule is proportional to its prior belief weight. For this alternative updating process, the main results for the centipede game (i.e., Theorems 1 and 2) still go through. However, we are not able to prove the same results for the repeated prisoner's dilemma and the chain store game.

Player i chooses the optimal rule L_{k^*} in round $t+1$ from the rule hierarchy $\{L_0, L_1, \ldots, L_S\}$ based on belief $B^i(t)$ in order to maximize expected payoffs. Let a_{ks} be the specified action of rule L_k at subgame s. Player i believes that action $a_{k's}$ will be chosen with probability $B^i_{k'}(t)$ by the opponent. Hence, the optimal rule chosen by player i is:

$$k^* = \text{argmax}_{k=1,\ldots,S} \quad \sum_{s=1}^{S} \left\{ \sum_{k'=1}^{S} B^i_{k'}(t) \cdot \pi_i(a_{ks}, a_{k's}) \right\}, \tag{III.5}$$

where $\pi_i(a_{ks}, a_{k's})$ is player i's payoff at subgame s if i chooses rule L_k and the opponent chooses $L_{k'}$ rule (cf. Camerer et al., 2004).

Note that we model learning across game rounds but not across stages within a game round. This is clearly an approximation. For a general game, a player could potentially update her belief about the opponent' rule within a round as the game unfolds. Specifically, a player may rule out a potential rule used by her opponent if an observed choice by the opponent at a particular stage is inconsistent with that rule. For example, in the 4-stage centipede game, if player B observes player A passing in stage I, player B will infer that player A's chosen rule must be level 3 or lower. Nevertheless, we believe that our model is a good starting point for 2 reasons. First, in the 3 games we study below (centipede game, repeated prisoner's dilemma, and chain-store game), the posterior belief at the end of a game round remains the same whether or not within-round learning is modeled explicitly. This is so because no new information is revealed once a player's opponent chooses an action that is contrary to the prior belief. For instance, in the centipede game, when a player is surprised by an opponent who takes earlier than expected, the game ends immediately and no additional information is revealed.[4] Second, within-round learning frequently generates prediction that is contrary to observed behavior. For example, in the centipede game, the second player who expects the first player to take immediately will be surprised if the latter passes. If we incorporate within-round learning, this will lead the second player to put more weights on the lower level rules. As a consequence, the second player will be more likely to pass, which generally runs

[4]Similarly, in the repeated Prisoner's dilemma, players' subsequent actions after a defection by some player does not contain new information as long as players choose from the rule hierarchy specified below. In the chain store game, once the incumbent shares the market, subsequent actions by entrants will provide no new information as long as they choose from the defined rule hierarchy.

counter to the observed data.

We need to determine player i's initial belief $B^i(0)$. We define $N^i(0)$ such that $N^i_k(0) = \beta$ for some k and 0 otherwise, where the parameter β captures the weight assigned to initial belief. In other words, player i places all the initial weight on a particular rule L_k and zero weight on all other rules. Different players have different guesses about others' rationality and hence place the initial weight on a different k. A player who places the initial weight on k will choose an initial rule of L_{k+1} in round 1. The heterogeneity of players' initial beliefs is captured by a Poisson distribution. The proportion of players who hold initial belief k is given by:

$$\phi(k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!}; \quad k = 0, \ldots, S. \tag{III.6}$$

For example, a $\phi(0)$ proportion of players believe that their opponents will play L_0 and thus choose L_1. Similarly, a $\phi(k)$ proportion of players believe that their opponents will play L_k and best-respond with L_{k+1}. All players best-respond given their beliefs in order to maximize their expected payoffs. Note that no players will choose L_0 under our model. That is, L_0 only occurs in the minds of the higher level players.[5]

The dynamic level-k model, similar to the cognitive hierarchy (CH) model, uses a Poisson distribution to capture player heterogeneity. The single parameter Poisson approach allows these models to be tractably used as a building block for theoretical analysis. However, the parsimony of a parametric approach comes with an empirical limitation. Specifically, the Poisson distribution often implies a non-negligible proportion of L_0 players in empirical applications of the CH model (see Camerer et al., 2004). In contrast, using a general discrete distribution, Costa-Gomes and Crawford (2006) and Crawford and Iriberri (2007a, 2007b) show that the estimated proportion of L_0 players is frequently zero. In agreement with the above observation, our proposed dynamic level-k model assumes that there are no L_0 players. Unlike the CH model, which captures heterogeneity in players' rules, the dynamic level-k model captures heterogeneity in players' beliefs of others' rules and allows players to best-respond to their beliefs. For example, $\phi(0)$ represents the proportion of L_0 players in

[5]In centipede games, this implies that the dynamic level-k does not admit any occurrence of Outcome 0, where players pass all the way. No model with only self-interested players can account for such behavior. Our model can easily be extended by incorporating social preferences to capture this kind of phenomenon.

the CH model, but it represents, in our model, the proportion of players who believe that their opponents play L_0 and thus choose L_1. Therefore, our proposed dynamic level-k model excludes L_0 players.[6] In this way, the dynamic level-k model retains the parsimony of the Poisson distribution while improving its empirical validity.

The dynamic level-k model is different from the static level-k and cognitive hierarchy models in 3 fundamental aspects. First, players in our model are not endowed with a specific thinking type. That is, players in our model are cognitively capable of choosing any rule but always choose the level that maximizes their expected payoff. In other words, players in our model are not constrained by reasoning ability. Second, players in our model may be aware of others who adopt higher level rules than themselves. In other words, a player who chooses L_k may recognize that there are others who choose L_{k+1} or higher but still prefers to choose L_k because there is a large majority of players who are L_{k-1} or below. On the other hand, the static level-k and CH models assume that players always believe they are the highest level thinkers (i.e., the opponents are always of a lower level rule). Third, unlike the static level-k and CH models, players in the dynamic level-k model may change their rules as they collect more information and update beliefs about others. Specifically, a player who interacts with opponents of higher level rules may advance to a higher rule. Similarly, a players who interacts with opponents of lower level rules may switch to a lower level rule in order to maximize their expected payoffs.

In summary, the dynamic level-k model has 2 parameters, β and λ. The parameter λ captures the degree of heterogeneity in initial beliefs and the parameter β captures the strength of the initial belief, which in turn determines players' sensitivity to game history. The 2-parameter dynamic level-k model nests several well-known special cases. When $\lambda = \infty$, it reduces to backward induction. If $\beta = \infty$, players have a stubborn prior and never respond to game history. This reduces our model to a variant of the static level-k model.[7] Consequently, we can empirically test whether these special cases are good approximations of behavior using

[6]It is possible to generalize this setup by introducing a segment of non-strategic players (level 0). Let π be the proportion of non-strategic players and $1 - \pi$ be the proportion of strategic players (level 1 or higher). As described above, the Poisson model can be used to describe the distribution of strategic players. Consequently, the dynamic level-k model is simply a special case with $\pi = 0$.

[7]Note that under the dynamic level-k model, while players may believe others are level-0, they themselves never play level-0 rules.

the standard generalized likelihood principle.

Below, we apply the dynamic level-k model to explain violations of backward induction in three canonical games (centipede game, finitely repeated prisoner's dilemma, and chain store game). In each case, we prove that the dynamic level-k model can account for the limited induction and time unraveling properties of the data.[8]

A. Centipede game

McKelvey and Palfrey (1992) study centipede games with an even number of stages (e.g., 4 and 6 stages). Hence, we focus on this class of games (ie., S is even). The above authors also assume there is a proportion of players who always pass in every stage. To avoid over-fitting, we assume that L_0 corresponds to the same strategy. That is, we choose $L_0 = \{P_1, P_2, \ldots, P_{S-1}, P_S\}$. (Note that odd-numbered components apply to Player A and even-numbered components apply to Player B.) A player who believes her opponent uses L_0 will maximize her payoff by adopting $L_1 = \{P_1, P_2, \ldots, P_{S-1}, T_S\}$. In general, L_k best-responds to L_{k-1} so that $L_k = \{P_1, \ldots, P_{S-k}, T_{S-k+1}, \ldots, T_S\}$. Therefore in a centipede game with S stages, $L_S = L_{S+1} = \ldots = L_\infty$. Put differently, all rules L_S or higher prescribe the same action at each stage as L_∞ (the backward induction rule). Consequently, we pool

[8] Besides these three games, alternating-offers bargaining game is also frequently used to show systematic violations of backward induction. Camerer et al. (1993) is one of the earliest experiments to show that subjects rarely look at some crucial payoff information that is needed to apply backward induction. Johnson et al. (2002) builds on this work by showing that these violations cannot be explained by social preferences since subjects make similar offers to robot players. In a related study, Binmore et al. (2002) identify 2 behavioral properties implied by the principle of backward induction and show that subjects frequently violate them. The first behavioral property is called truncation consistency, which suggests that behavior should be invariant to replacing a subgame with its equilibrium payoff. In the centipede game, this implies that behaviors in a 4-stage game should be identical to behavior in the first four stages in a 6-stage game. The experimental data from McKelvey and Palfrey (1992) strongly rejected this property. The limited induction property predicts this violation and hence can help explain why truncation consistency is violated in alternating-offers bargaining games. The second behavioral property identified is called subgame consistency, which implies that behavior should be invariant to whether the same game is played independently or part of a larger game. The dynamic level-k model cannot explain violation of subgame consistency in alternating-offers bargaining games. There was, however, no evidence to suggest that subgame consistency is violated in experimental centipede games (McKelvey and Palfrey, 1992).

all these higher level rules together and collectively call them L_S.

Under the dynamic level-k belief model, players choose rules from the rule hierarchy L_k ($k = 0, 1, 2, 3, \ldots$). Let $L^i(t)$ be the rule of Player i (where $i = A, B$) in time t. Then $\delta(L^A(0), L^B(0), G)$ is the deviation from backward induction L_∞ in game G at time 0. For example, if Players A and B both play L_2, the deviation is $\frac{1}{2}$ in a 4-stage and $\frac{2}{3}$ in a 6-stage centipede game. Let G_4 and G_6 be the 4-stage and 6-stage games respectively. Then we have the following theorem:

Theorem 1 *In the centipede game, the dynamic level-k model implies that the limited induction property is satisfied. Formally, if the initial distribution of beliefs is the same (i.e., λ is the same) in both games G_S and $G_{S'}$ with $S < S'$, then the expected deviation from backward induction in G_S, $\delta(L^A(0), L^B(0), G_S)$, is smaller than in $G_{S'}$, $\delta(L^A(0), L^B(0), G_{S'})$.*

Proof: See appendix.

Theorem 1 suggests that the dynamic level-k model gives rise to a smaller deviation from backward induction in a game with a smaller number of stages. This result is consistent with the data presented in Figure 2. The Appendix gives the detailed proof but the basic idea of the proof is outlined here. Given any rule combination $L^A(0)$ and $L^B(0)$ for the players, let $K^A(0) = 2 \cdot \lfloor \frac{L^A(0)}{2} \rfloor$ and $K^B(0) = 2 \cdot \lceil \frac{L^B(0)}{2} \rceil - 1$. The outcome is identical in both games (counting from the last stage) (see Figure 1). Specifically, the game outcome is $z = \max\{K^A(0), K^B(0)\}$ and the number of actions that are inconsistent with backward induction is $S - z$. As a consequence, G_6 has a larger deviation than G_4 (see equation (I.1)). Since the initial distribution of beliefs is the same in both games (i.e., having the same λ), the initial expected deviation must be higher in G_6.

Figure 3 shows that the expected deviation from backward induction becomes smaller over time. There is a question of whether this trend will persist and eventually unravel to backward induction outcome. The following theorem formally states that this is indeed the case for the dynamic level-k model.[9]

[9] Theorem 2 shows that all players will choose $L_S = L_\infty$ as $t \to \infty$. However, during the transient phase, it is possible for players to switch from L_k to $L_{k'}$ ($k' < k$) if they repeatedly encounter opponents of lower level rules. This phenomenon is occasionally observed in the data and can be accommodated by our model.

Theorem 2 *In a centipede game with S stages, the dynamic level-k model implies that the time unraveling property is satisfied. Formally, the deviation from backward induction $\delta(L^A(t), L^B(t), G)$ converges to zero and all players will choose $L_S = L_\infty$ as $t \to \infty$. That is, players will eventually take in every stage.*

Proof: See appendix.

Theorem 2 states that the dynamic level-k model satisfies the time unraveling property. The basic idea of the proof is outlined here. No players choose L_0. As a consequence, L_1 players will learn that other players are L_1 or higher. Using our notation, this means that $B_0^i(t)$ will decline over time and the speed of decline depends on the initial belief weight β. For a specific β, there is a corresponding number of rounds after which all L_1 players will move up to L_2 or higher. No players will then choose L_0 and L_1. In the same way, L_2 players will learn that other players are L_2 or higher and will eventually move to L_3 or higher. Consequently, we will see a 'domino' phenomenon whereby lower level players will successively disappear from the population. In this regard, players believe that others become more sophisticated over time and correspondingly do so themselves. In the limit, all players converge to L_S (and the learning process ceases).

The proof also reveals an interesting insight. The number of rounds it takes for L_k to disappear is increasing in k. For example, when $\beta = 1$, it takes 6 rounds for L_0 to disappear and another 42 rounds for L_1 to disappear. In general, for an initial belief weight β, it takes a total of $(7^k - 1)\beta$ rounds for rule L_k to disappear. Note that each higher level rule takes an exponentially longer time to be eliminated from the population.[10] This result suggests that time unraveling occurs rather slowly in the centipede game. In the experiments conducted by McKelvey and Palfrey (1992), there are only 10 game rounds and no substantial learning is observed. In Section IV, we fit the dynamic level-k model and show that only L_1 disappears in their data set.

[10] In our model, the cumulative rule counts $N_k^i(t)$ do not decay over time. If decay is allowed, that is, $N_k^i(t) = \delta \cdot N_k^i(t-1) + I(k,t) \cdot 1$, then unraveling can occur at a much faster rate. For example, if $\delta = 0$, it takes only one round for each successively lower level rule to disappear.

B. Finitely Repeated Prisoner's Dilemma

In the finitely repeated prisoner's dilemma, backward induction prescribes that all players should defect in every repetition (see Figure 4 for a game with 10 repetitions). This is because defection is a dominant strategy in every repetition. To explain cooperative behavior, Kreps et al. (1982) assume there is a proportion of players who always adopt the Tit-For-Tat (TFT) rule, which starts with cooperation and reciprocates by choosing the opponent's action in the previous round. To avoid over-fitting, we choose L_0 to be TFT. A player who believes her opponent uses L_0 will play TFT until the very last round when it is optimal to defect. We denote this rule by $L_1 = \{TFT, D\}$. Note that L_1 is better than TFT because defection yields a payoff of 5 in the last round while cooperation yields a payoff of 3; in contrast, $\{TFT, D, D\}$ is inferior to $\{TFT, D\}$ because the former yields a payoff of 5+1=6 in the last 2 rounds whereas the latter yields a payoff of 3+5=8 (see Figure 4). In general, L_k best-responds to L_{k-1} so that $L_k = \{TFT, \overbrace{D, \ldots, D}^{k}\}$.[11] Therefore, in a repeated prisoner's dilemma game with S repetitions, $L_S = L_{S+1} = \ldots = L_\infty$. Put differently, all rules L_S or higher prescribe the same actions as backward induction. Consequently, we pool all these higher level rules together and collectively call them L_S.

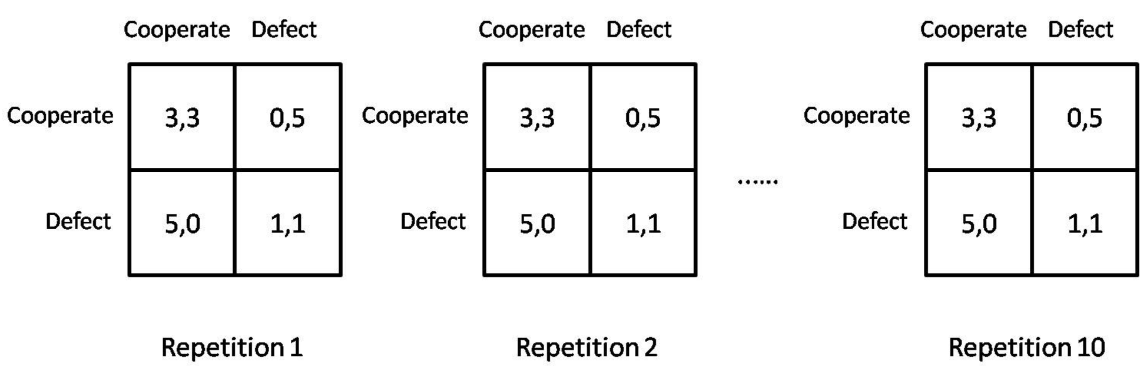

Figure 4: Finitely repeated prisoner's dilemma games (10 repetitions).

Theorem 3 *In the finitely repeated prisoner's dilemma, the dynamic level-k model implies*

[11] Choosing cooperation initially followed by defection in the last k repetitions is also a best-response to L_{k-1}. However, this simpler rule will perform poorly when played against a higher level rule. Since players are fully aware of the entire rule hierarchy and may believe that others can use a higher level rule, it is in their best interest to adopt a more robust TFT based rule.

that the limited induction property is satisfied. Formally, if the initial distribution of beliefs is the same (i.e., λ is the same), then the expected deviation from backward induction is smaller in a game with a smaller number of repetitions.

Proof: See appendix.

The basic idea of the proof is similar to the one in the centipede game. For any combination of initial rules $L^A(0)$ and $L^B(0)$, let $z = \max\{L^A(0), L^B(0)\}$. Then defection begins after $S - z$ repetitions, and both players will choose defection in the remaining $z - 1$ repetitions. As a consequence, a repeated prisoner's dilemma with a larger number of repetitions will have a longer string of sustained cooperation before defection begins. Thus the expected deviation of backward induction is smaller for a game with fewer repetitions.

Theorem 4 *In the finitely repeated prisoner's dilemma with S repetitions, the dynamic level-k model implies that the time unraveling property is satisfied. Formally, the deviation from backward induction converges to zero and all players will choose $L_S = L_\infty$ as $t \to \infty$. That is, players will eventually defect in every repetition.*

Proof: See appendix.

Similar to before, no players choose L_0 (i.e., TFT), so all players will defect in the last stage. L_1 players will learn over time that other players are L_1 or higher and eventually move to L_2 or higher (i.e., L_1 disappears). After this transition, all players will defect in the second last repetition. The same logic continues and eventually all players will defect in every single repetition. Like before, the number of rounds for L_k to disappear is increasing in k. In general, for an initial belief weight β, it takes a total of $(2^k - 1)\beta$ rounds for rule L_k to disappear. Note that the rate of convergence is faster for repeated prisoner's dilemma than centipede game. In general, the rate of convergence depends on the payoff structure of the game.

C. Chain Store Paradox

In the chain store game, an incumbent faces potential competition from numerous entrants in separate locations and must interact with each one of them sequentially (Selten, 1978).

Entrants in each location can either choose to enter the market or stay out. In Figure 5, there are 10 entrants. If an entrant chooses to stay out, the chain store receives a payoff of \$5 m and the entrant receives a payoff of \$1 m. If the entrant enters the market, the chain store must then choose to either stage a price war or share the market with the entrant. If the chain store fights, both players receive \$0 m. If the chain store shares, both players receive \$2 m.

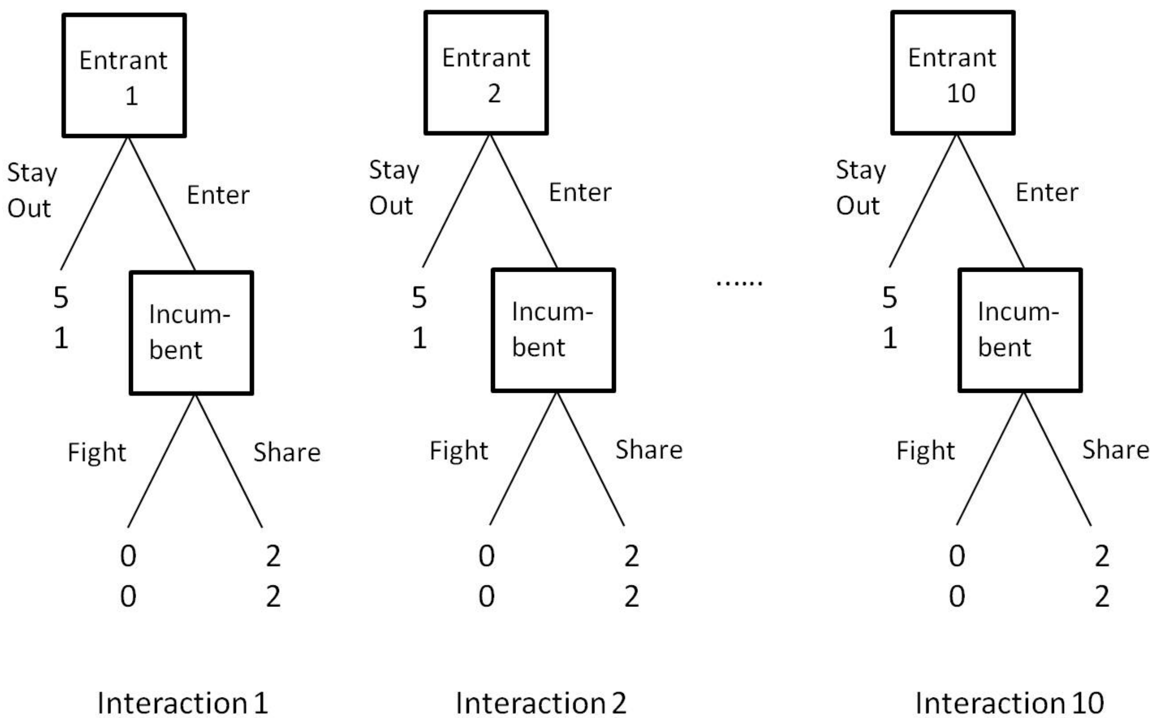

Figure 5: Chain store game with 10 entrants over 10 interactions.

Focusing on the very last entrant, the chain store will choose to share the market with the entrant if it enters. Repeating this logic, backward induction predicts that each entrant will enter at its respective location and the chain store will share the market with each one of them. This seems implausible because introspection suggests that the chain store is likely to fight at least in some of the markets and not all entrants will choose to enter in their locations. Specifically, Selten (1978) wrote:

> *The disturbing disagreement between plausible game behavior and game theoretical reasoning constitutes the chain store paradox.*

To explain entry and sharing, Jung et al. (1994) and Camerer et al. (2002) assume that there is a probability that the incumbent is a fighter type, who always fights upon entry. To avoid over-fitting, we assume that the L_0^I incumbent will always fight conditional on entrance. We assume that L_1^E entrants who believe the incumbent uses L_0^I will maximize their payoff by adopting the so-called grim trigger rule (GTR). The GTR rule prescribes that an entrant contemplating entry will choose to stay out unless the chain store is observed to have shared the market in the past. Given the sequential nature of the game, L_1^I incumbent will best-respond to L_1^E entrants by always fighting until the very last interaction when it is optimal to share the market condition on entrance, i.e., $\{F, F, \ldots, F, S\}$.

In general, L_k^E entrants best-respond to a L_{k-1}^I incumbent and this implies that $L_k^E = \{GTR, \overbrace{E, \ldots, E}^{k-1}\}$. Similarly, a L_k^I incumbent best-responds to L_{k-1}^E entrants and this implies that $L_k^I = \{F, F, F, \overbrace{S, \ldots, S}^{k}\}$. Therefore, in a chain store game with S interactions, $L_{S+1}^E = \ldots = L_\infty^E$ and $L_S^I = L_{S+1}^I = \ldots = L_\infty^I$. Put differently, all rules with level S or higher prescribe the same actions as the rule of backward induction. Consequently, we pool all these higher level rules together and collectively call them L_{S+1}^E and L_S^I.

Note that we have defined two separate rule hierarchies, one for each role, because the incumbent and entrants have a different strategy space. Further, in this game, each interaction is a sequential game where the incumbent chooses only after observing the action of the entrant. Consequently, our rule hierarchies are defined in a zig-zag structure where L_k^E entrants best-respond to L_{k-1}^I incumbent and L_k^I incumbent always best-responds to L_k^E entrants. Note that the zig-zag structure is for notational convenience and it does not change the main results.

Theorem 5 *In the chain store game, the dynamic level-k model implies that the limited induction property is satisfied. Formally, if the initial distribution of beliefs is the same (i.e., λ is the same) and all entrants have the same initial belief (i.e. they assign β to the same k), then the expected deviation from backward induction is smaller in a game with a smaller number of interactions.*

Proof: See appendix.

In this theorem, the basic idea is similar to before. Let $L^I(0)$ and $L^E(0)$ be the initial rules for incumbent and all entrants respectively. For any combination of rule $L^I(0)$ and $L^E(0)$, the outcome in the last $z = \max\{L^I(0), L^E(0) - 1\}$ interactions is the same regardless of the total number of interactions. As a consequence, a chain store game with a larger number of repetitions will have a longer string of entrants staying out and the incumbent fighting before entrants begin to enter. Thus the expected deviation of backward induction is smaller for a game with fewer interactions.

Theorem 6 *In the chain store game with S interactions, the dynamic level-k model implies that the time unraveling property is satisfied. Formally, the deviation from backward induction converges to zero, that is, the incumbent will choose $L^I_S = L^I_\infty$ and entrants will choose $L^E_{S+1} = L^E_\infty$ as $t \to \infty$. That is, every entrant will eventually enter and incumbent will always share in every interaction.*

Proof: See appendix.

The idea of the proof is similar to that in the previous two games. No incumbent will always fight. Given this fact, L^E_1 entrants will learn that the incumbent is L^I_1 or higher and eventually will share in the last interaction. This implies that L^E_1 will disappear. After this happens, the L^I_1 incumbent will learn that entrants are L^E_2 or higher. As a consequence, L^I_1 will eventually disappear. The same logic continues and eventually all entrants will enter and the incumbent will share the market in every single interaction. This is consistent with backward induction.

D. Properties of Level-0 Rule

The above rule hierarchies are generated from iterated best responses to a specific level-0 rule. To avoid over-fitting, we have used the level-0 rules from the existing literature. For instance, following McKelvey and Palfrey (1992), we use a decision rule that passes in every stage as level-0 in the centipede game. Similarly, as in Kreps et al. (1982), the Tit-For-Tat rule is used as level-0 in the repeated prisoner's dilemma. Finally, in chain store game, we define always fighting as level-0 for the incumbent (following Jung et al., 1994 and Camerer et al. 2002).

There are 2 ways to formally determine the level-0 rule. First, we can structurally estimate the level-0 rule using the data. One can parameterize the space of possible decision rules and empirically determine the best-fitting level-0 rule. Second, one can develop a theory for choosing a level-0 rule. We explore the latter approach below. Further research can provide a more comprehensive analysis.

A good level-0 rule may satisfy the following 2 attractive properties:

1. Maximize group payoff: A level-0 player always chooses a decision rule that if others do the same will lead to the largest total payoff for the group. For instance, in the centipede game, if every player chooses a decision rule that passes in every stage, the group will achieve the largest possible payoff.

2. Protect individual payoff: While maximizing group payoff, a level-0 player also ensures that the chosen decision rule is robust against continued exploitation by others. For instance, in the repeated prisoner's dilemma, a level-0 who attempts to maximize group payoff by always cooperating may be exploited by an individual who always chooses to defect. A robust rule like Tit-For-Tat will prevent this kind of exploitation and promote cooperative behavior.

The above two properties do not define a unique level-0 rule in all games. However, they do narrow down the possible candidates for consideration. For the above 3 games that we consider, the chosen level-0 rules satisfy the two properties.

IV. An Empirical Application to Centipede Game

A. Dynamic Level-k Model

We use the dynamic level-k model to explain violations of backward induction in experimental centipede games conducted by McKelvey and Palfrey (1992). The authors ran experiments using students subjects from Caltech (2 sessions) and Pasadena Community College (PCC) (4 sessions). In each subject pool, half the sessions were run on the 4-stage game and the other half on the 6-stage game. Each experimental session consisted of 18 or 20 subjects and each subject played the game in the same role either 9 or 10 times. The random matching protocol was such that each player was matched with another player only once.

Contrary to backward induction, players did not always take immediately in both 4-stage and 6-stage games. In fact, a large majority passed in the first stage. For instance, 94% of the Caltech subjects passed in the first stage in 4-stage games (see Tables 1a-b). The distribution of game outcomes is unimodal with the mode occurring at the intermediate outcomes (Outcomes 2 and 3 in 4-stage games, Outcomes 3 and 4 in 6-stage games). These results present a considerable challenge to backward induction.

Tables 1a-b also suggest that Caltech subjects take one stage earlier than PCC subjects in both 4 and 6-stage games. Specifically, the modal outcome is Outcome 3 in 4-stage games and Outcome 4 in 6-stage games in Caltech subject pool while it is Outcome 2 in 4-stage games and Outcome 3 in 6-stage games in PCC subject pool. These results suggest that the two subject pools exhibit different levels of sophistication. We estimate the dynamic level-k model using the data from both subject pools with a goal of explaining their main features.

Outcome	4	3	2	1	0
Caltech (N=100)	0.06	0.43	0.28	0.14	0.09
PCC (N=181)	0.08	0.31	0.42	0.16	0.03

Table 1a: A Comparison of Outcomes in 4-stage Games Between Caltech and PCC subjects

Outcome	6	5	4	3	2	1	0
Caltech (N=100)	0.02	0.09	0.39	0.28	0.20	0.01	0.01
PCC (N=181)	0	0.05	0.09	0.44	0.28	0.12	0.02

Table 1b: A Comparison of Outcomes in 6-stage Games Between Caltech and PCC subjects

We conduct a conservative test for our model by specifying the model prediction for all game rounds without using the real-time data.[12] As a consequence, the prediction of the dynamic

[12]It is possible to use an alternative approach where real-time data is used to adjust the model prediction. This adjustment will allow heterogeneity across both initial rules and time. This estimation approach may

level-k model for every game round is completely parameterized by λ and β. Conditional on λ and β, the model generates players' choice probabilities for each rule L_k for all game rounds. These choice probabilities in turn determine the distribution of outcomes over time. Let $P_t^i(k|\lambda,\beta)$ be the model's predicted probability that player i will choose a rule L_k in round t. In the first round, $P_1^i(k|\lambda,\beta) = \phi(k-1)$. In each subsequent round, to generate $P_t^i(k|\lambda,\beta)$, we adopt a large sample approach by considering a large population of players.[13] At the end of round t, players update their prior belief $B^i(t-1)$ based on their opponents' chosen rule, which is drawn from $P_t^i(k|\lambda,\beta)$ using a random matching protocol. This updating process gives rise to $B^i(t)$. Each player then chooses a rule L_k that maximizes her expected payoff. If two rules yield the same expected payoffs, players are assumed to always choose the simpler rule (i.e. a lower level rule).[14] In this way, we can then compute the choice probabilities $P_{t+1}^i(k|\lambda,\beta)$ in round $t+1$.

Let $O(t)$ be the outcome of a game in round t. For example, in the 6-stage game, $O(t) = 6$ if Player A takes immediately and $O(t) = 0$ if everyone passes to the end. Let $P(O(t))$ be the probability of observing outcome $O(t)$. One can compute the probability of observing a data point $O(t)$ conditional on λ and β. The probability is

$$P(O(t)) = \left[\sum_k \sum_{k'} P_t^A(k|\lambda,\beta) \cdot P_t^B(k'|\lambda,\beta) \cdot I\left(H(k,k'), O(t)\right)\right], \qquad \text{(IV.1)}$$

where $I(H(k,k'), O(t))$ is 1 if $H(k,k') = O(t)$ and 0 otherwise. We define $H(k,k') = \max\left\{2 \cdot \lfloor \frac{k}{2} \rfloor, 2 \cdot \lceil \frac{k'}{2} \rceil - 1\right\}$.

The dynamic level-k model makes the following sharp predictions in the centipede game depending on the value of the β parameter:

provide a better fit but runs the risk of over-fitting. We tried this method and found that subjects moved up and down the rule hierarchy more frequently than the data would suggest. Consequently we chose not to adopt this approach.

[13]If we do not use the large sample approach, the model prediction for all game rounds can still be generated. However, we can no longer assume that the distribution of choices is conditionally independent from the distribution of choices in previous rounds. Therefore, we need to consider all possible combinations of each player's initial beliefs and how these beliefs evolve over time. This gives rise to a combinatorial problem as the number of rounds increases and makes the estimation computationally intractable.

[14]For example, in the centipede game, for player B, $L_1 = L_2 = \{-, P, -, T\}$. We assume that player B always chooses L_1 over L_2.

1. Players' choice probabilities $P_t^i(k|\lambda,\beta)$ do not change at every round. In fact, they change at most twice within 10 game rounds.

2. For $\beta \geq \frac{5}{3}$, $P_t^i(k|\lambda,\beta) = \phi(k-1)$ remains fixed over all 10 game rounds. For $\frac{5}{3} > \beta \geq \frac{1}{6}$, $P_t^i(k|\lambda,\beta)$ changes only once within 10 game rounds. For $\beta < \frac{1}{6}$, $P_t^i(k|\lambda,\beta)$ changes twice within 10 game rounds.

3. The model starts by predicting that outcome 0 (i.e., players pass all the way) occurs with zero probability. At every change of the choice probabilities, the next higher outcome (outcome 1 at the first change and outcome 2 at the second change) will also occur with zero probability.

The dynamic level-k predicts that some game outcomes will occur with zero probability. To facilitate empirical estimation, we need to incorporate an error structure. To avoid specification bias, we use the simplest possible error structure (see Crawford and Iriberri 2007a). We assume a probability $(1-\epsilon)$ that the data matches our model prediction and a probability ϵ that the observed outcome is uniformly distributed over all possible outcomes. Hence, the likelihood function is given by:

$$L = \prod_t \left[(1-\epsilon) \cdot P(O(t)) + \epsilon \cdot \frac{1}{S+1} \right]. \tag{IV.2}$$

We fit the dynamic level-k model to the data using maximum likelihood estimation. We separately estimate the dynamic level-k model for Caltech and PCC subject pools because of the apparent differences in their behaviors (see Tables 1a-b). However, we use the same set of parameters to fit both the 4-stage and 6-stage games.

Table 2 shows the estimation results for Caltech subjects. As shown, the estimated parameters are $\hat{\lambda} = 1.32, \hat{\beta} = 0.25, \hat{\epsilon} = 0.29$, and the log-likelihood is -305.8. The estimated $\hat{\lambda} = 1.32$ suggests that initially, 27% of the players are level 1, 35% are level 2, 23% are level 3, and the remaining 15% are level 4 or higher. The estimated $\hat{\beta}$ lies between $\frac{1}{6}$ and $\frac{5}{3}$, suggesting that players' choice probabilities change only once. The estimated $\hat{\beta} = 0.25$ suggests a small initial belief weight, which implies that time unraveling can begin to occur very quickly and in fact after only 2 rounds.[15]

[15] Any value of β between $\frac{1}{6}$ and $\frac{1}{3}$ will predict a change after round 2 and we simply report the midpoint of these 2 values.

Parameter	Backward Induction	Naive Belief	Static Level k	Full Model
λ	∞	0	1.76	1.32
β	∞	0.08	∞	0.25
ϵ	1.00	0.97	0.21	0.29
Log Likelihood	-357.1	-355.0	-312.3	-305.8

Table 2: Maximum Likelihood Estimates of Dynamic Level-k Model and Its Special Cases (Caltech Subjects)

Table 3a-b compare the model prediction and the actual data for the first 2 and last 8 rounds in the 4-stage and 6-stage games for Caltech subjects. The best fitted dynamic level-k model makes two predictions. First, in both games, Outcome 0 (i.e., both players pass all the way) should not occur in all rounds. Second, the model prescribes that unraveling occurs, and as a consequence Outcome 1 should not occur after round 2. These predictions were roughly consistent with the data in the following ways:

1. In both games, Outcome 0 occurs infrequently. For example, Outcome 0 occurs less than 1% of the time in 6-stage game. The model however fails to capture a nontrivial proportion of occurrence of Outcome 0 in the 4-stage game (even though the proportion declines from 15% to 8% after round 2).

2. In 4-stage game, the proportions of Outcomes 0 and 1 decline after 2 rounds. For example, the proportion of Outcome 1 decreases from 25% to 11%. In 6-stage game, Outcomes 0 and 1 rarely occur after round 2 (about 1% of the time).

Outcome	4	3	2	1	0
Rounds 1-2 Data (N=20)	0	0.35	0.25	0.25	0.15
Prediction ($\epsilon = 0$)	0.15	0.32	0.36	0.17	0
Rounds 3-10 Data (N=80)	0.08	0.45	0.29	0.11	0.08
Prediction ($\epsilon = 0$)	0.27	0.52	0.21	0	0

Table 3a: A Comparison of Data and Dynamic Level-k Model Prediction in 4-stage games (Caltech Subjects)

Outcome	6	5	4	3	2	1	0
Rounds 1-2 Data (N=20)	0	0	0.30	0.25	0.45	0	0
Prediction ($\epsilon = 0$)	0.01	0.04	0.13	0.29	0.36	0.17	0
Rounds 3-10 Data (N=80)	0.03	0.11	0.41	0.29	0.14	0.01	0.01
Prediction ($\epsilon = 0$)	0.01	0.06	0.24	0.47	0.21	0	0

Table 3b: A Comparison of Data and Dynamic Level-k Model Prediction in 6-stage games (Caltech Subjects)

Finally, the error rate is around 29% which is not negligible. We believe the value is attributed to the sharp prediction of the dynamic level-k model that both outcome 0 and 1 should not occur after round 2.

We also fit three nested cases of the dynamic level-k model. They are all rejected by the likelihood ratio test. The backward induction model (i.e., $\lambda = \beta = \infty$) yields a log-likelihood of -357.1, which is strongly rejected in favor of the full model ($\chi^2 = 102.7$). The naive belief model (i.e., $\lambda = 0$) (which assumes that all players believe that their opponents will always pass) yields an estimated $\hat{\beta} = 0.08$ and a log-likelihood of -355.0. The model is again strongly rejected with a $\chi^2 = 98.5$. The static level-k model corresponds to $\beta = \infty$. Since players' initial belief will persist throughout the game plays, they will always choose the same rule L_k across rounds. This restriction yields a parameter estimate of $\hat{\lambda} = 1.76$ and provides a reasonable fit with a log-likelihood of -312.3. However, the static model is also rejected by the likelihood ratio test ($\chi^2 = 13.1, p < 0.01$).

Given the MLE estimates of $\hat{\lambda} = 1.32, \hat{\beta} = 0.25$, we can generate the dynamic level-k model's predicted frequencies for each of the outcomes. Figure 6 shows the actual and predicted frequencies of each outcome. The top panel shows the results for the 4-stage game and the bottom panel shows the results for the 6-stage game. Backward induction predicts that only Outcome 4 in 4-stage game and Outcome 6 in 6-stage game can occur (i.e., Player A takes immediately). This backward induction prediction is strongly rejected by the data. As shown, the dynamic-k model does a reasonable job in capturing the unimodal distribution of the outcomes. In addition, the dynamic level-k model is able to capture the two most frequently occurring outcomes. Specifically, the dynamic level-k model correctly predicts that the two most frequently played outcomes are 2 and 3 in 4-stage game and 3 and 4 in 6-stage game.

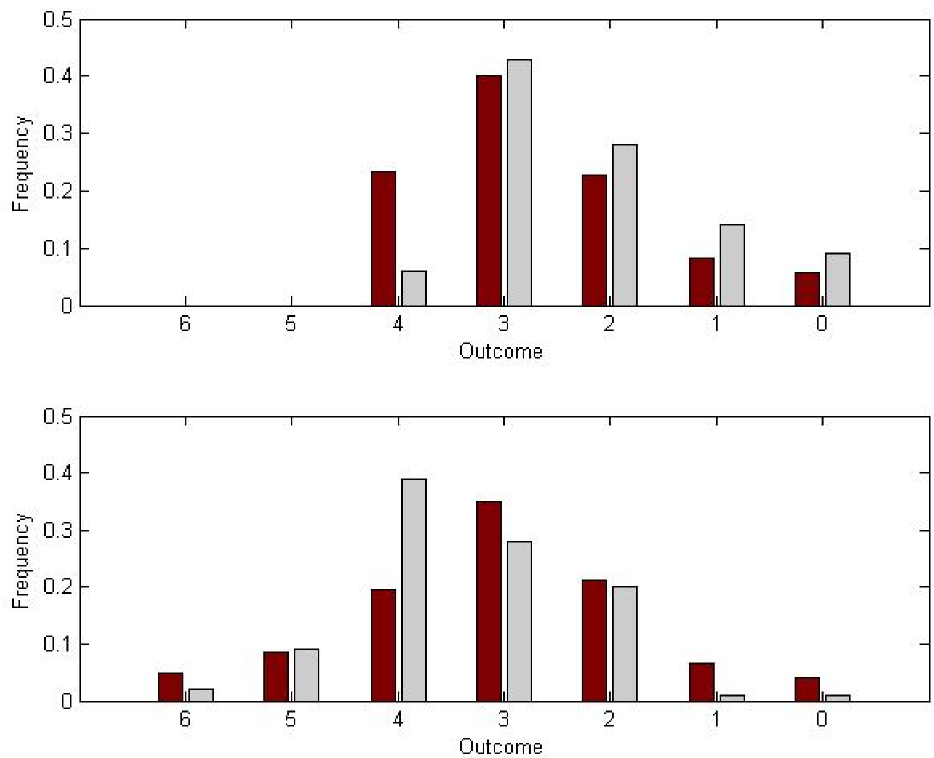

Figure 6: Dynamic Level-k Model Fit (dark bars) and Data (light bars) (Caltech subjects).

We also fit our model to the data obtained from PCC subjects. Table 4 shows the estimation results. The dynamic level-k model yields parameter estimates of $\hat{\lambda} = 1.30$ and $\hat{\beta} = 1.67$. This estimated initial belief weight $\hat{\beta} > \frac{5}{3}$ suggests that unraveling never occurs (i.e., there is no change in the choice probabilities). Therefore, for PCC subjects, the dynamic level-k model delivers identical predictions as the static level-k model (i.e., $\beta = \infty$), which also yields an estimated $\hat{\lambda} = 1.30$. Both the static and dynamic level-k models give a log-likelihood score of -514.8. In contrast, the backward induction prediction ($\lambda = \beta = \infty$) gives a log-likelihood score of -646.5 and the naive belief model ($\lambda = 0$) gives a log-likelihood score of -644.3, so both models are strongly rejected by the likelihood ratio test.

Parameter	Backward Induction	Naive Belief	Static Level k	Full Model
λ	∞	0	1.30	1.30
β	∞	1.67	∞	1.67
ϵ	1.00	0.99	0.08	0.08
Log Likelihood	-646.5	-644.3	-514.7	-514.7

Table 4: Maximum Likelihood Estimates of Dynamic Level-k Model and Its Special Cases (PCC Subjects)

Since the dynamic level-k model predicts that there is no unraveling for PCC subjects, it is interesting to examine the actual data to see whether this is indeed the case. Tables 5a-b show the predictions of the dynamic level-k model as well as the actual data. The data suggests that PCC subjects exhibit a slower process of time unraveling compared to Caltech subjects (cf. Tables 3a-b). For example, after 2 rounds of play in the 6-stage game, Outcomes 0 and 1 still occur with more than 10% probability in PCC subjects while the same outcomes occur with only 2% chance in Caltech subjects. In fact, for both 4-stage and 6-stage games, the t-test provides evidence of a significant difference between the mean outcomes in the first 2 rounds and last 8 rounds for Caltech subjects. However, the same test is inconclusive for PCC subjects. These results show that within the 10 game rounds in the data, time unraveling has occurred for Caltech subjects but not for PCC subjects.

Outcome	4	3	2	1	0
Rounds 1-2 Data (N=38)	0.08	0.24	0.42	0.24	0.03
Prediction ($\epsilon = 0$)	0.26	0.39	0.27	0.08	0
Rounds 3-10 Data (N=143)	0.08	0.34	0.42	0.14	0.03
Prediction ($\epsilon = 0$)	0.26	0.39	0.27	0.08	0

Table 5a: A Comparison of Data and Dynamic Level-k Model Prediction in 4-stage games (PCC Subjects)

Outcome	6	5	4	3	2	1	0
Rounds 1-2 Data (N=38)	0	0.13	0.05	0.32	0.32	0.16	0.03
Prediction ($\epsilon = 0$)	0.03	0.10	0.20	0.31	0.27	0.08	0
Rounds 3-10 Data (N=143)	0	0.03	0.10	0.48	0.27	0.10	0.01
Prediction ($\epsilon = 0$)	0.03	0.10	0.20	0.31	0.27	0.08	0

Table 5b: A Comparison of Data and Dynamic Level-k Model Prediction in 6-stage games (PCC Subjects)

Finally, given the MLE estimates of $\hat{\lambda} = 1.30, \hat{\beta} = 1.67$ for PCC subjects, we can generate the dynamic level-k model's predicted frequencies for each of the outcomes. Figure 7 shows the actual and predicted frequencies of each outcome. The top panel shows the results for the 4-stage game and the bottom panel shows the results for the 6-stage game. The plot shows that the dynamic level-k model provides a reasonably good fit of the data.

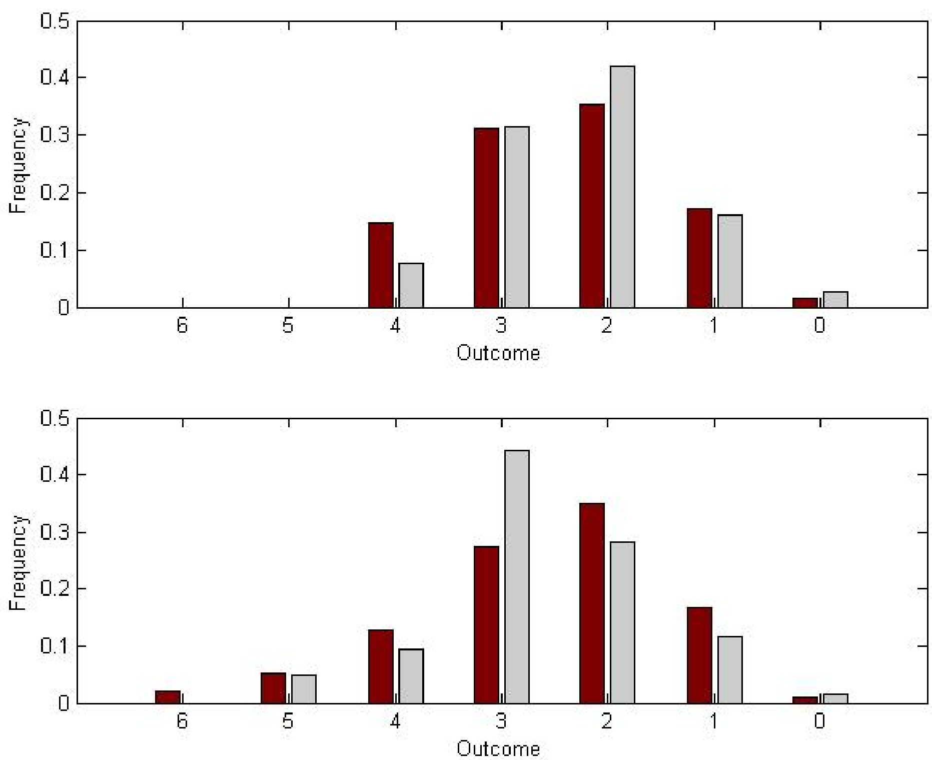

Figure 7: Dynamic Level-k Model Fit (dark bars) and Data (light bars) for PCC subjects.

B. Reputation-based Model

As indicated before, backward induction does not admit passing behavior in the centipede game. To explain why players pass, one can transform the game into a game of incomplete information by introducing some uncertainty over the player type (Kreps et al., 1982). If there is a small fraction θ of players that always pass (the so-called altruists), then even the self-interested players may find it in their best interest to pass initially. However, this reputation-based model generates the following empirical implications that run counter to the observed data (see McKelvey and Palfrey, 1992):

1. When θ is small ($0 \leq \theta \leq \frac{1}{49}$), the model will not correctly predict the most frequently occurring outcomes. For instance, the model predicts the modal outcome to be Outcome 4 in 4-stage game while the data suggests it should be Outcome 2. Similarly, the model predicts the two most frequently occurring outcomes to be Outcomes 4 and 5 in 6-stage game while the data suggests they should be Outcomes 2 and 3.

2. When θ is intermediate ($\frac{1}{49} < \theta \leq \frac{1}{7}$), the model will not admit Outcome 4 and 4-6 in 4-stage and 6-stage respectively.

3. When θ is large ($\theta > \frac{1}{7}$), the model predicts players either pass all way (Outcome 0) or take at the last stage (Outcome 1), which is contrary to the data.

The above discussion suggests that there is not a common θ value that will explain the unimodal distribution of game outcomes in both the 4-stage and 6-stage games well.

We estimate the reputation-based model using the centipede game data. Like before, we assume that there is a probability ϵ that the observed outcome will be evenly distributed over all possible outcomes. The best fitted model yields $\hat{\theta} = 0.050, \hat{\epsilon} = 0.62$ for Caltech subjects and $\hat{\theta} = 0.075, \hat{\epsilon} = 0.31$ for PCC subjects. The log-likelihood scores are -329.8 and -518.8 respectively, suggesting that the model did a worse job in fitting the data compared to the dynamic level-k model.

C. Models of Social Preferences

Could the unimodal distribution of game outcomes in centipede game be attributed to social preferences? The answer is no. To see this we apply the inequity aversion model of Fehr and Schmidt (1999). Let x_i and x_{-i} denote the material payoffs of player i and opponent $-i$ respectively. Then, the payoff of player i, $U_i(x_i, x_{-i})$, is given by:

$$U_i(x_i, x_{-i}) = x_i - \alpha \cdot [x_i - x_{-i}]^+ - \beta \cdot [x_{-i} - x_i]^+ \tag{IV.3}$$

where $0 < \alpha < 1$ captures a player's aversion to being ahead and $\beta > 0$ captures a player's aversion to being behind. The parameter α is between 0 and 1 because a player who is ahead will not give up a \$1 to benefit her opponent less than \$1. One can then solve the game by backward induction given these revised payoff functions.

Theorem 7 *In the centipede game, social preferences lead to either taking immediately or passing all the way in both the 4-stage and 6-stage games. Specifically, social preferences predicts that if* $3\cdot\alpha - 6\cdot\beta < 2$*, players will take immediately; otherwise they will pass all the way.*

Proof: See Appendix.

The above theorem states that a model of social preferences admits only either Outcome 0 or 4 in 4-stage game and either Outcome 0 or 6 in 6-stage game. The occurrence of either outcome depends on the fairness parameters α and β. If players are sufficiently averse to being behind (high β), they will take immediately, which is consistent to backward induction. On the other hand, if players are sufficiently averse to being ahead (high α), they will pass all the way. Both predictions are inconsistent with the unimodal distribution of actual game outcomes.[16]

We estimate the social preference model using the centipede game data. Note that the model fit is identical to the backward induction model. If $3 \cdot \alpha - 6 \cdot \beta \leq 2$, the model reduces to backward induction, so the best fit yields a log-likelihood of -357.1 for Caltech subjects and -646.5 for PCC subjects. If $3 \cdot \alpha - 6 \cdot \beta \geq 2$, the model predicts that players pass all the way, so the best fit yields again a log-likelihood of -357.1 and -646.5. Hence the model performs poorly compared to the dynamic level-k model.

[16] Fey et al. (1996) study a constant-sum version of the centipede game in which there are initially two equal piles of cash. Similar to the regular centipede game, each player takes turn to either take or pass. The player who takes will receive the larger pile of cash. Each time a player passes, a fraction of the smaller pile is transferred to the larger pile, so the division of the money becomes more unequal over stages. This design helps to control for social preferences. They show that about half the subjects still pass in the first stage and hence violate backward induction. This result suggests that violations of backward induction cannot be fully accounted for by social preferences.

V. Conclusions

In economic experiments, backward induction is frequently violated. We develop a dynamic level-k model to explain two systematic violations of backward induction. First, players tend to deviate more from backward induction in games with a greater number of stages or subgames. Our model captures this limited induction by allowing players to have heterogeneous initial beliefs about others' rationality and hence adopt different rules from a rule hierarchy. Second, players move closer to backward induction over time. Our model captures this time unraveling by allowing players to update beliefs of their opponents' rules and hence adjust their own rules over time. We show that this adjustment process leads to convergence towards backward induction.

We have applied the dynamic level-k model to 3 canonical games (centipede game, finitely repeated prisoner's dilemma, and chain store game) that demonstrate the limitations of backward induction. In all three games, we prove that the limited induction and time unraveling properties hold. Limited induction holds because the same rule always has a higher deviation from backward induction in a game with more stages. Time unraveling occurs because the dynamic level-k model implies a domino effect over time as lower level rules are successively eliminated.

We fit our model to experimental centipede games of McKelvey and Palfrey (1992). Our estimation results show that the dynamic level-k model captures the unimodal distribution of game outcomes in both 4-stage and 6-stage games reasonably well. Special cases including backward induction and the static level-k model are strongly rejected by the data. Interestingly, Caltech subjects, who are arguably more sophisticated, tend to learn faster than PCC subjects.

We rule out 2 alternative explanations for the observed phenomena in the centipede game. First, we show that the reputation-based model of Kreps et al. (1982) cannot capture the unimodal distribution of game outcomes in the manner consistent with the data. A limitation of the reputation-based model is that there is no common fraction of the altruistic players that will simultaneously account for the degree of passing in both the 4-stage and 6-stage games. Second, we show that social preferences in the form of inequity aversion (Fehr and Schmidt 1999) lead to players either taking immediately or passing all the way. Consequently, it cannot explain the unimodal distribution of game outcomes in both games. Also, a model of social preferences cannot capture both the limited induction and time unraveling properties.

Our model is a generalization of backward induction. Since the model converges to backward induction in the limit, it can be conceptualized as a tracing procedure for backward induction, hence providing a dynamic foundation for backward induction. Framing it this way, the violations of backward induction in the experiments are simply "transient" behaviors and our model explicitly characterizes that trajectory.

References

[1] Aumann, R. (1995) 'Backward Induction and Common Knowledge of Rationality,' Games and Economic Behavior, 8: 6-19.

[2] Axelrod, R. (1984) The Evolution of Cooperation. Basic Books, Inc. Publishers. New York.

[3] Bernheim, B. D. (1984) 'Rationalizable Strategic Behavior', Econometrica, 52(4): 1007-1028.

[4] Binmore, K., J. McCarthy, G. Ponti, L. Samuelson and A. Shaked (2002) 'A Backward Induction Experiment,' Journal of Economic Theory, 104: 48-88.

[5] Camerer, C. and T-H. Ho (1999) 'Experience-Weighted Attraction Learning in Normal Form Games,' Econometrica, 67: 837-874.

[6] Camerer, C., E. J. Johnson, S. Sen S. and T. Rymon (1993) 'Cognition and Framing in Sequential Bargaining for Gains and Losses,' in *Frontiers of Game Theory* (K. Binmore, A. Kirman, and P. Tani, eds.) MIT Press, Cambridge, MA.

[7] Camerer, C.F., T-H. Ho, J-K. Chong (2004) 'A Cognitive Hierarchy Model of Games,' Quarterly Journal of Economics, 119(3): 861-898.

[8] Cooper, R., D. V. DeJong, R. Forsythe, and T. Ross (1996) 'Cooperation without Reputation: Experimental Evidence from Prisoner's Dilemma Games,' Games and Economic Behavior, 12: 187-218.

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

[10] Costa-Gomes, M. and V. Crawford (2006) 'Cognition and Behavior in Two-Person Guessing Games: An Experimental Study,' American Economic Review, 96: 1737-1768.

[11] Costa-Gomes, M., V. Crawford, and N. Iriberri (2009) 'Comparing Models of Strategic Thinking in Van Huyck, Battalio, and Beil's Coordination Games,' Journal of the European Economic Association, 7: 377-387.

[12] Crawford, V. P. (2003) 'Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions,' American Economic Review 93: 133-149.

[13] Crawford, V. and N. Iriberri (2007a) 'Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games,' American Economic Review, 97: 1731-1750.

[14] Crawford, V. P. and N. Iriberri (2007b) 'Level-k Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?,' Econometrica, 75: 1721-1770.

[15] Fehr, E. and K.M. Schmidt (1999) 'A Theory of Fairness, Competition, and Cooperation,' Quarterly Journal of Economics, 114(3): 81768.

[16] Fey, M., R. D. McKelvey, and T. R. Palfrey (1996) 'An Experimental Study of Constant-Sum Centipede Games,' International Journal of Game Theory, 25(3): 269-87.

[17] Fudenberg, D. and D. Levine (1998) 'Theory of Learning in Games,' Cambridge, MIT Press.

[18] Jung, Y. J., J. H. Kagel, and D. Levin (1994) 'On the Existence of Predatory Pricing: An Experimental Study of Reputation and Entry Deterrence in the Chain-Store Game,' The RAND Journal of Economics, 25(1): 72-93.

[19] Ho, T-H., C. Camerer, and K. Weigelt (1998) 'Iterated Dominance and Iterated Best Response in Experimental P-Beauty Contests,' The American Economic Review, 88: 947-969.

[20] Ho, T-H., C. Camerer, C. and J-K. Chong (2007) 'Self-tuning Experience-Weighted Attraction Learning in Games,' Journal of Economic Theory, 133(1): 177-198.

[21] Johnson, E., C. Camerer, S. Sen, and T. Rymon (2002) 'Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining,' Journal of Economic Theory, 104:16-47.

[22] Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982) 'Rational Cooperation in the Finitely Repeated Prisoners' Dilemma,' Journal Economic Theory, 27: 245-252.

[23] Levitt, S. D., J. A. List, and S. E. Sadoff (2009) 'Checkmate: Exploring Backward Induction Among Chess Players,' Working Paper, University of Chicago Economics Department.

[24] McKelvey, R. D. and T. R. Palfrey (1992) 'An Experimental Study of the Centipede Game,' Econometrica, 60(4): 803-836.

[25] Palacios-Huerta I. and O. Volijc (forthcoming) 'Field Centipedes,' The American Economic Review.

[26] Pearce, D. G. (1984) 'Rationalizable Strategic Behavior and the Problem of Perfection,' Econometrica, 52(4): 1029-1050.

[27] Reny, P. J. (1992) 'Backward Induction, Normal Form Perfection and Explicable Equilibria,' Econometrica, 60(3): 627-649.

[28] Rosenthal, R. (1981), 'Games of Perfect Information, Predatory Pricing, and the Chain Store,' Journal of Economic Theory, 25: 92-100.

[29] Selten, R. (1978) 'The Chain Store Paradox,' Theory and Decision, 9: 127-159.

[30] Stahl D. and P. Wilson (1995) 'On Players Models of Other Players - Theory and Experimental Evidence,' Games and Economic Behavior, 10: 213-254.

[31] Stahl, D. (1996) 'Boundedly Rational Rule Learning in a Guessing Game,' Games and Economic Behavior, 16: 303-330.

[32] Stahl, D. (2000) 'Rule Learning in Symmetric Normal-Form Games: Theory and Evidence,' Games and Economic Behavior, 32: 105-138.

Appendix A: Proofs

Proof of Theorem 1 Consider 2 centipede games with s and S subgames, where $s < S$. Suppose that the proportion of players who choose each rule level is the same in both games. Let π_k denote the proportion of players who choose rule L_k. Let $E\delta(S)$ denote the expected deviation in a game with S subgames.

Suppose Player A adopts rule L_{k_A} and Player B adopts rule L_{k_B} in a game with S stages. Then, the deviation is

$$\delta(L_{k_A}, L_{k_B}, S) = \frac{1}{S}\left[S - (\min(\lfloor k_A/2 \rfloor, S/2) + \min(\lceil k_B/2 \rceil, S/2))\right],$$

which is weakly increasing in S for every k_A and k_B.

Now, comparing the expected deviations in both games, we have

$$E\delta(S) - E\delta(s) = \sum_{k_A=1}^{S} \sum_{k_B=1}^{S} \pi_{k_A}\pi_{k_B}[\delta(L_{k_A}, L_{k_B}, S) - \delta(L_{k_A}, L_{k_B}, s)] \geq 0,$$

which gives the required result.

■

Proof of Theorem 2 Consider any player in the centipede game. For concreteness, we focus on the 4-stage game (see Figure 1) but the proof proceeds in the same way for any number of stages. Suppose this player holds beliefs $B(t) = (b_0, b_1, \ldots, b_4)$ at time t, where b_k denotes this player's belief of the probability that the opponent will use rule L_k. Let $\mathbf{b}_j^k \equiv \sum_{i=j}^{k} b_i$. Given these beliefs, the player chooses a best-response rule to maximize expected payoffs.

Let V_s^i denote Player i's expected payoff, when there are s stages remaining, from following an optimal strategy henceforth. In the last stage, Player B always takes regardless of beliefs since taking yields 32 a payoff of but passing yields only a payoff of 16. Therefore, $V_1^B = 32$. In the second-last stage, Player A's expected payoff is 16 from taking and $64b_0 + 8(1 - b_0)$ from passing (because after passing, Player B will pass again if and only if L_0 rule is used, which occurs with probability b_0). Thus, $V_2^A = \max\{16, 64b_0 + 8(1 - b_0)\}$ and Player A takes if and only if $b_0 \leq 1/7$. Similarly, in the third-last stage, Player B's expected payoff is 8 from taking and $V_1^B \mathbf{b}_0^1 + 4(1 - \mathbf{b}_0^1) = 32\mathbf{b}_0^1 + 4(1 - \mathbf{b}_0^1)$ from passing, so $V_3^B = \max\{8, 32\mathbf{b}_0^1 + 4(1 - \mathbf{b}_0^1)\}$ and Player B takes if and only if $\mathbf{b}_0^1 \leq 1/7$. Finally, in the first stage of the 4-stage game, Player A's expected payoff is 4 from taking and $V_2^A \mathbf{b}_0^2 + 2(1 - \mathbf{b}_0^2)$ from passing. Note that if $\mathbf{b}_0^2 \leq 1/7$, we must have $b_0 \leq \mathbf{b}_0^2 \leq 1/7$ and $V_2^A = 16$, so $V_2^A \mathbf{b}_0^2 + 2(1 - \mathbf{b}_0^2) \leq 4$ and Player A takes; but if $\mathbf{b}_0^2 \geq 1/7$, we must have $V_2^A \mathbf{b}_0^2 + 2(1 - \mathbf{b}_0^2) \geq 16\mathbf{b}_0^2 + 2(1 - \mathbf{b}_0^2) \geq 4$ so Player A passes. Therefore, Player A takes if and only if $\mathbf{b}_0^2 \leq 1/7$. Applying this logic, observe that in the s-th last stage of the centipede game (where $s \geq 2$), the player will take if and only if $\mathbf{b}_0^{s-2} \leq 1/7$. Therefore, given beliefs $B(t) = (b_0, b_1, \ldots, b_4)$, the best response rule L_{k^*} must be such that k^* is the largest integer satisfying $\mathbf{b}_0^{k^*-2} \leq 1/7$.

In the model, each player starts at $t = 0$ with a belief weight of β on some k and adds a unit weight at each round t. Note that no player will choose L_0, so the weights will be added to levels $k \geq 1$ in each round. Therefore, at every $t \geq 6\beta$, we must have $b_0 \leq 1/7$. In other words, after round $t = 6\beta$, no player will choose L_0 or L_1, so the weights will be added to levels $k \geq 2$ in subsequent rounds. Similarly, after another $6 \cdot 7\beta$ rounds, at every $t \geq (6 + 6 \cdot 7)\beta$, we must have $\mathbf{b}_0^1 \leq 1/7$. Continuing with this logic, note that at every $t \geq (7^k - 1)\beta$, we must have $\mathbf{b}_0^{k-1} \leq 1/7$. Therefore, in a S-stage game, at every $t \geq (7^{S-1} - 1)\beta$, we have $\mathbf{b}_0^{S-2} \leq 1/7$, so all players use $L_S = L_\infty$.

■

Proof of Theorem 3 Consider 2 repeated prisoner's dilemma games with s and S repetitions, where $\Delta \equiv S - s > 0$. Suppose that the proportion of players who choose each rule level is the same in both games. Let π_k denote the proportion of players who choose rule L_k. Let $E\delta(S)$ denote the expected deviation in a game with S repetitions.

Suppose the players adopt rules of levels k_1 and k_2. Denote $\overline{k} = \max(k_1, k_2)$ and $\underline{k} = \min(k_1, k_2)$. Then, the deviation is

$$\delta(L_{k_1}, L_{k_2}, S) = \frac{1}{S}\left[S - \min(\overline{k}, S) + \frac{\mathbf{1}_{\{k_1 \neq k_2\}}}{2}\right]$$

because mutual cooperation is sustained in the first $S - \min(\overline{k}, S)$ repetitions and one of the players continues to cooperate in the next round if $k_1 \neq k_2$.

Next, we consider the above deviation in several special cases. Denote the sets $A = \{0, 1, \ldots, s-1\}, B = \{s, s+1, \ldots, S-1\}, C = \{S, S+1, \ldots\}$. If $k_1 = k_2 \equiv k$, we have the following cases:

$$
\begin{aligned}
k \in A \quad &\Rightarrow \quad \delta(S) = \frac{s-k+\Delta}{s+\Delta}, \quad \delta(s) = \frac{s-k}{s} \\
k \in B \quad &\Rightarrow \quad \delta(S) = \frac{s-k+\Delta}{s+\Delta}, \quad \delta(s) = 0 \\
k \in C \quad &\Rightarrow \quad \delta(S) = \delta(s) = 0
\end{aligned}
$$

If $k_1 \neq k_2$, we have the following cases:

$$
\begin{aligned}
\underline{k} \in A, \overline{k} \in A \quad &\Rightarrow \quad \delta(S) = \frac{s-\overline{k}+1/2+\Delta}{s+\Delta}, \quad \delta(s) = \frac{s-\overline{k}+1/2}{s} \\
\underline{k} \in A, \overline{k} \in B \quad &\Rightarrow \quad \delta(S) = \frac{S-\overline{k}+1/2}{S} \geq \frac{1/2}{S}, \quad \delta(s) = \frac{1/2}{s} \\
\underline{k} \in A, \overline{k} \in C \quad &\Rightarrow \quad \delta(S) = \frac{1/2}{S}, \quad \delta(s) = \frac{1/2}{s} \\
\underline{k} \in B \cup C \quad &\Rightarrow \quad \delta(S) \geq \delta(s) = 0
\end{aligned}
$$

Finally, we compute the expected deviations in both games. We obtain:

$$
\begin{aligned}
E\delta(S) - E\delta(s) \quad &= \quad \sum_{k_1=1}^{S}\sum_{k_2=1}^{S} \pi_{k_1}\pi_{k_2}[\delta(L_{k_1}, L_{k_2}, S) - \delta(L_{k_1}, L_{k_2}, s)] \\
&\geq \quad \sum_{k_1,k_2 \in A} \pi_{k_1}\pi_{k_2}[\delta(L_{k_1}, L_{k_2}, S) - \delta(L_{k_1}, L_{k_2}, s)] \\
&\quad + \sum_{\underline{k} \in A, \overline{k} \in B \cup C} \pi_{k_1}\pi_{k_2}[\delta(L_{k_1}, L_{k_2}, S) - \delta(L_{k_1}, L_{k_2}, s)] \\
&\geq \quad \sum_{k_1,k_2 \in A} \pi_{k_1}\pi_{k_2}\left[\frac{s-\overline{k}+\mathbf{1}_{\{k_1 \neq k_2\}}/2+\Delta}{s+\Delta} - \frac{s-\overline{k}+\mathbf{1}_{\{k_1 \neq k_2\}}/2}{s}\right] \\
&\quad + \sum_{\underline{k} \in A, \overline{k} \in B \cup C} \pi_{k_1}\pi_{k_2}\left[\frac{1/2}{S} - \frac{1/2}{s}\right] \\
&\geq \quad \sum_{k_1,k_2 \in A} \pi_{k_1}\pi_{k_2}\left[\frac{s-\overline{k}+\mathbf{1}_{\{k_1 \neq k_2\}}/2+\Delta}{s+\Delta} - \frac{s-\overline{k}+\mathbf{1}_{\{k_1 \neq k_2\}}/2}{s}\right] \\
&\quad + \sum_{k_1,k_2 \in A} \pi_{k_1}\pi_{k_2}\left[\frac{1}{S} - \frac{1}{s}\right] \\
&= \quad \sum_{k_1,k_2 \in A} \pi_{k_1}\pi_{k_2}\left[\frac{s-\overline{k}+\mathbf{1}_{\{k_1 \neq k_2\}}/2+\Delta+1}{s+\Delta} - \frac{s-\overline{k}+\mathbf{1}_{\{k_1 \neq k_2\}}/2+1}{s}\right] \\
&\geq \quad 0.
\end{aligned}
$$

The second last inequality holds because $Prob\{\underline{k} \in A, \overline{k} \in B \cup C\} \geq 2 \cdot Prob\{k_1, k_2 \in A\}$, and the last inequality holds because $\frac{s-\overline{k}+\mathbf{1}_{\{k_1 \neq k_2\}}/2+1}{s} \leq 1$ whenever $k_1, k_2 \in A$. Therefore, the expected deviation is higher in a game with more repetitions.

■

Proof of Theorem 4 Consider any player in the repeated prisoner's dilemma with S repetitions. Suppose this player holds beliefs $B(t) = (b_0, b_1, \ldots, b_S)$ at time t, where b_k denotes this player's belief of the probability that the opponent will use rule L_k. Let $\mathbf{b}_j^k \equiv \sum_{i=j}^k b_i$. Given these beliefs, the player chooses a best-response rule to maximize expected payoffs.

Let V_s denote the player's expected continuation payoff from following an optimal strategy henceforth, when there are s stages remaining and mutual cooperation was sustained for all previous $S - s$ stages.

In the last repetition, the player always defects (since this is a dominant strategy), so $V_1 = 5b_0 + 1 \cdot (1 - b_0)$, since there is a probability of b_0 that the opponent will use rule L_0 and cooperate in the last repetition. In general, in the s-th last repetition, the player's expected continuation payoff from defecting is $D_s \equiv (5 + (s-1))\mathbf{b}_0^{s-1} + s(1 - \mathbf{b}_0^{s-1})$ because the opponent cooperates in this repetition with probability $\mathbf{b}_0^{s-1}$ and the payoff in all subsequent repetitions will be 1; similarly, the player's expected continuation payoff from cooperating is $C_s \equiv (3 + V_{s-1})\mathbf{b}_0^{s-1} + (s-1)(1 - \mathbf{b}_0^{s-1})$. Therefore, for every value of s, we obtain $V_s = \max\{D_s, C_s\} = \max\{(5+(s-1))\mathbf{b}_0^{s-1} + s(1-\mathbf{b}_0^{s-1}), (3+V_{s-1})\mathbf{b}_0^{s-1} + (s-1)(1-\mathbf{b}_0^{s-1})\}$.

Next, we show that the intermediate result that $V_{s+1} - V_s \geq 1$ for every $s \geq 1$. To see this, first, note that $V_2 - 1 \geq V_1$ because $V_2 - 1 \geq D_2 - 1 = [5\mathbf{b}_0^1 + (1 - \mathbf{b}_0^1)] \geq V_1$, so the statement above holds for $s = 1$. Second, note that if $V_s - V_{s-1} \geq 1$, it then follows that

$$\begin{aligned}
C_{s+1} - C_s &= [(3 + V_s)\mathbf{b}_0^s + s(1 - \mathbf{b}_0^s)] - [(3 + V_{s-1})\mathbf{b}_0^{s-1} + (s-1)(1 - \mathbf{b}_0^{s-1})] \\
&\geq [(3 + V_s)\mathbf{b}_0^{s-1} + s(1 - \mathbf{b}_0^{s-1})] - [(3 + V_{s-1})\mathbf{b}_0^{s-1} + (s-1)(1 - \mathbf{b}_0^{s-1})] \\
&= [(V_s - V_{s-1})\mathbf{b}_0^{s-1} + (1 - \mathbf{b}_0^{s-1})] \\
&\geq 1,
\end{aligned}$$

where the first inequality above holds because $3 + V_s \geq s$. The above result, together with the fact that $D_{s+1} - D_s \geq 1$, implies that $V_{s+1} - V_s \geq 1$. Therefore, by induction, we have $V_{s+1} - V_s \geq 1$ for every $s \geq 1$. Note that this result implies that $V_s - s$ is increasing in s.

When there are s repetitions remaining, the player cooperates if and only if

$$\begin{aligned} C_s &\geq D_s \\ (3+V_{s-1})\mathbf{b}_0^{s-1} + (s-1)(1-\mathbf{b}_0^{s-1}) &\geq (5+(s-1))\mathbf{b}_0^{s-1} + s(1-\mathbf{b}_0^{s-1}) \\ V_{s-1}\mathbf{b}_0^{s-1} &\geq (2+(s-1))\mathbf{b}_0^{s-1} + (1-\mathbf{b}_0^{s-1}) \\ (V_{s-1}-s)\mathbf{b}_0^{s-1} &\geq 1. \end{aligned}$$

Since both $V_s - s$ and $\mathbf{b}_0^s$ are increasing in s, the player's optimal rule given beliefs $B(t) = (b_0, b_1, \ldots, b_S)$ must be L_{k^*}, where k^*+1 is the smallest integer s satisfying $(V_{s-1}-s)\mathbf{b}_0^{s-1} \geq 1$.

In the model, each player starts at $t = 0$ with a belief weight of β on some k and adds a unit weight at each round t. Note that no player will choose L_0, so the weights will be added to levels $k \geq 1$ in each round. Therefore, at every $t \geq \beta$, we must have $b_0 \leq 1/2$, which implies $V_1 = 5b_0 + 1 \cdot (1-b_0) \leq 3$ and hence $(V_1 - 2)\mathbf{b}_0^1 \leq V_1 - 2 \leq 1$, so $C_2 \leq D_2$. In other words, after round $t = \beta$, no player will choose L_0 or L_1, so the weights will be added to levels $k \geq 2$ in subsequent rounds. Similarly, after another 2β rounds, at every $t \geq 3\beta$, we must have $\mathbf{b}_0^1 \leq 1/2$. This then implies that $V_2 = D_2 = 6\mathbf{b}_0^1 + 2(1-\mathbf{b}_0^1) \leq 4$ and hence $(V_2 - 3)\mathbf{b}_0^2 \leq V_2 - 3 \leq 1$, so $C_3 \leq D_3$ and the player will choose rules with level 3 or higher. Continuing with this logic, note that at every $t \geq (2^k - 1)\beta$, we must have $\mathbf{b}_0^{k-1} \leq 1/2$ and the player will choose only rules of level $k+1$ or higher. Therefore, in a prisoner's dilemma with S repetitions, at every $t \geq (2^{S-1}-1)\beta$, we have $\mathbf{b}_0^{S-2} \leq 1/2$, so all players use $L_S = L_\infty$.

■

Proof of Theorem 5 Consider 2 chain store games with s and S interactions, where $s < S$. In each game, assume that the entrants maintain the same beliefs about the incumbent, since they observe the same game history. In addition, suppose that the proportion of players who choose each rule level is the same in both games. Let π_k denote the proportion of players who choose rule L_k. Let $E\delta(S)$ denote the expected deviation in a game with S subgames.

Suppose the incumbent uses rule level k_I and the entrants use rule level k_E. Then the deviation is

$$\delta(L_{k_I}, L_{k_E}, S) = \frac{1}{S}\left\{[S-k_E]^+ + \frac{[\min(k_E,S) - \min(k_I,S)]^+}{2}\right\}.$$

This is because given the above rules, the entrant chooses to stay out in the first $[S-k_E]^+$ interactions, and if $k_E > k_I$ the incumbent will fight and the entrant will enter for the next $[\min(k_E,S) - \min(k_I,S)]$ rounds.

Consider two cases. If $k_E \leq k_I$, the above reduces to

$$\delta(L_{k_I}, L_{k_E}, S) = \begin{cases} 0, & S \leq k_E, \\ \frac{1}{S}(S - k_E), & S > k_E, \end{cases}$$

which is increasing in S. Similarly, if $k_E > k_I$, we have

$$\delta(L_{k_I}, L_{k_E}, S) = \begin{cases} 0, & S \leq k_I, \\ \frac{1}{S}[(S - k_I)/2], & k_I < S \leq k_E, \\ \frac{1}{S}\left[(S - k_E) + (k_E - k_I)/2\right], & S > k_E, \end{cases}$$

which is also increasing in S. Therefore, comparing the expected deviations in both games, we have

$$E\delta(S) - E\delta(s) = \sum_{k_I=1}^{S} \sum_{k_E=1}^{S} \pi_{k_I} \pi_{k_E} [\delta(L_{k_I}, L_{k_E}, S) - \delta(L_{k_I}, L_{k_E}, s)] \geq 0,$$

which gives the required result.

■

Proof of Theorem 6 Consider any player in the chain store game with S interactions. Suppose this player holds beliefs $B(t) = (b_0, b_1, \ldots, b_S)$ at time t, where b_k denotes this player's belief of the probability that the opponent will use rule L_k. Let $\mathbf{b}_j^k \equiv \sum_{i=j}^{k} b_i$. Given these beliefs, the player chooses a best-response rule to maximize expected payoffs.

We first show that the incumbent will eventually share in every interaction.

Let V_s^E denote the incumbent's expected continuation payoff from following an optimal strategy henceforth, when: (i) there are s stages remaining, (ii) the incumbent has not shared during any of the previous $S-s$ interactions, and (iii) the entrant has just entered. Similarly, V_s^O denote the incumbent's expected continuation payoff from following an optimal strategy henceforth, when (i) and (ii) above hold but the incumbent has not observed the decision of the current entrant.

In the last interaction, the incumbent always shares if the entrant enters, so $V_1^E = 2$. Further, $V_1^O = 5b_0 + V_1^E(1 - b_0) = 5b_0 + 2(1 - b_0)$, since there is a probability of b_0 that the entrant will use rule L_0 and stay out in the last interaction. In general, in the s-th last repetition, after the entrant enters, the incumbent's expected continuation payoff is $2s$ from sharing and V_{s-1}^O from fighting. Therefore, we have $V_s^E = \max\{2s, V_{s-1}^O\}$ and the incumbent fights with s interactions remaining if and only if $V_{s-1}^O \geq 2s$. It also follows that $V_s^O = (5 + V_{s-1}^O)\mathbf{b}_0^{s-1} + V_s^E(1 - \mathbf{b}_0^{s-1})$.

Next, we show that $V_{s-1}^O \geq 2s$ implies $V_s^O \geq 2(s+1)$ for every $s \geq 2$. We first check this for $s = 2$: when $V_1^O - 4 = b_0 - 2(1 - b_0) \geq 0$, we have

$$\begin{aligned} V_2^O - 6 &= [3 + (V_1^O - 4)]\mathbf{b}_0^1 + \{[V_1^O - 4]^+ - 2\}(1 - \mathbf{b}_0^1) \\ &\geq (V_1^O - 4) + 3\mathbf{b}_0^1 - 2(1 - \mathbf{b}_0^1) \\ &\geq (V_1^O - 4) + 3b_0 - 2(1 - b_0) \geq 0, \end{aligned}$$

as stated. In general, suppose $V_{s-1}^O \geq 2s$. Further, suppose $V_{t-1}^O \geq 2t$ holds for $t = s-1, s-2, \ldots, s-r+1$ but not for $t = s-r$. Since the above inequality holds for $t = s-r+1$ but not for $t = s-r$, we know that $V_{s-r-1}^O - 2(s-r) < 0$ but

$$V_{s-r}^O - 2(s-r+1) = [3 + (V_{s-r-1}^O - 2(s-r))]\mathbf{b}_0^{s-r-1} + \{[V_{s-r-1}^O - 2(s-r)]^+ - 2\}(1 - \mathbf{b}_0^{s-r-1}) \geq 0,$$

so we must have $3\mathbf{b}_0^{s-r-1} - 2(1 - \mathbf{b}_0^{s-r-1}) \geq 0$ and hence $3\mathbf{b}_0^{s-1} - 2(1 - \mathbf{b}_0^{s-1}) \geq 0$. Note that if $r = s-1$ above (i.e., $V_{t-1}^O \geq 2t$ holds for all t from 2 to s), then we also have $3\mathbf{b}_0^{s-1} - 2(1 - \mathbf{b}_0^{s-1}) \geq 3b_0 - 2(1 - b_0) \geq b_0 - 2(1 - b_0) = V_1^O - 4 \geq 0$. Therefore, we obtain

$$\begin{aligned} V_s^O - (s+1) &= [3 + (V_{s-1}^O - s)]\mathbf{b}_0^{s-1} + \{[V_{s-1}^O - s]^+ - 2\}(1 - \mathbf{b}_0^{s-1}) \\ &\geq (V_{s-1}^O - s) + 3\mathbf{b}_0^{s-1} - 2(1 - \mathbf{b}_0^{s-1}) \geq 0, \end{aligned}$$

as stated.

When there are s interactions remaining, the incumbent fights if and only if $V_{s-1}^O \geq 2s$. From the result in the previous paragraph, we know that the incumbent will fight in all previous interactions if the same applies to the current interaction. Therefore, given beliefs $B(t) = (b_0, b_1, \ldots, b_S)$, the incumbent's best-response rule must be L_{k^*}, where $k^* + 1$ is the smallest integer s satisfying $V_{s-1}^O \geq 2s$.

In the model, each incumbent starts at $t = 0$ with a belief weight of β on some k and adds a unit weight at each round t. Note that no incumbent will choose L_0, so the weights will be added to levels $k \geq 1$ in each round. Therefore, after a finite number of rounds, we must have $b_0 \leq 2/3$, which implies $V_1^O - 4 = b_0 - 2 \cdot (1 - b_0) \leq 0$. In other words, after a finite number of rounds, the incumbent will never choose L_0 or L_1, so the weights will be added to levels $k \geq 2$ in all subsequent rounds. Similarly, after another finite number of rounds, we must have $\mathbf{b}_0^1 \leq 2/5$. This then implies that $V_2^O \leq 3\mathbf{b}_0^1 - 2(1 - \mathbf{b}_0^1) \leq 0$, so the incumbent will choose rules with level 3 or higher subsequently. Continuing with this logic, note that for every k, after a finite number of rounds, we must have $\mathbf{b}_0^{k-1} \leq 2/5$ and the incumbent will choose only rules of level $k+1$ or higher. Therefore, in a chain store game with S interactions, after a finite number of rounds, we have $\mathbf{b}_0^{S-2} \leq 2/5$, so the incumbent will use $L_S = L_\infty$ subsequently.

Finally, we show that the entrant will eventually share in every interaction. Consider an entrant that is followed by s subsequent entrants. This entrant's expected payoff is 1 from staying out and $2(1-\mathbf{b}_0^s)$ from entering. Thus, this entrant will enter if and only if $\mathbf{b}_0^s \leq 1/2$. Following the same logic above, note that for every s, we will eventually have $\mathbf{b}_0^s \leq 1/2$ after a finite number of rounds. Therefore, in a chain store game with S interactions, the entrant will eventually enter at every interaction.

■

Proof of Theorem 7 We will prove the result for a 4-stage game. The proof for a 6-stage game proceeds similarly. First, consider the case where $3\alpha - 6\beta \leq 2$. In the last stage, the payoffs to Players A and B respectively are (8,32) if Player B takes and (64,16) if Player B passes. The corresponding utilities to Player B are $32 - 24\alpha$ and $16 - 48\beta$ respectively. Since the condition $3\alpha - 6\beta \leq 2$ holds, Player B prefers to take. Reasoning backwards, in the second last stage, the payoffs to both players are (16,4) if Player A takes and (8,32) if Player A passes (since Player B will then take in the last stage). The corresponding utilities to Player A are $16 - 12\alpha$ and $8 - 24\beta$, so by the same condition above, Player A takes. Using the same reasoning, players will take in every stage.

Next, consider the other case where $3\alpha - 6\beta \geq 2$. By the calculations above, we know that Player B will now pass in the last stage. Now, in the second last stage, the payoffs to Players A and B are (16,4) if Player A takes and (64,16) if Player A passes (since Player B will also pass in the last stage). The utilities are $16 - 12\alpha$ and $64 - 48\alpha$. Since $\alpha \leq 1$, the latter is greater and thus Player A passes. Proceeding backwards, in the third last stage (i.e., Stage II), Player B's payoff from taking is $8 - 6\alpha$, which is less than his payoff $32 - 24\alpha$ from taking in the last stage, because $\alpha \leq 1$. Since Player B prefers passing in the last stage, the same applies in this stage as all players will pass subsequently. Using the same logic, players will pass in every stage.

■