Deep and shallow thinking in the long run*

Heinrich Nax† and Jonathan Newton‡

December 10, 2021

Abstract

Humans differ in their strategic reasoning abilities and in beliefs about others' strategic reasoning abilities. Studying such cognitive hierarchies has produced new insights regarding equilibrium analysis in economics. This paper investigates the effect of cognitive hierarchies on long run behavior. Despite short run behavior being highly sensitive to variation in strategic reasoning abilities, this variation is not replicated in the long run. In particular, when generalized risk dominant strategy profiles exist, they emerge in the long run independently of the strategic reasoning abilities of players. These abilities may be arbitrarily low or high, heterogeneous across players and evolve over time.

Keywords: bounded rationality, level-k thinking, evolution.

JEL Codes: C73, D81, D90.

1. Introduction

> *"Coordination, when it occurs, is an almost accidental (though statistically predictable) by-product of non-equilibrium thinking"*
>
> – Vincent Crawford (2007)

There is evidence that humans sometimes reason iteratively to predict the behavior of others and that the depth of such reasoning can vary according to person and situation (see Crawford, 2019, for a survey). Apart from some notable exceptions (e.g. Sáez-Martı and

*The authors gratefully acknowledge the comments of seminar audiences at Ca' Foscari University of Venice, University of Padua, University of Oregon, Bar-Ilan University and the Asia Meetings of the Econometric Society, as well as Xiaohong Chen, Vince Crawford, Ryota Iijima, Rosemarie Nagel, María Sáez Martí, Larry Samuelson, the editor and three anonymous referees. Nax's work was supported by a Swiss National Science Foundation Eccellenza Grant. Newton's work was supported by KAKENHI B funding from the Japan Society for the Promotion of Science (Grant 21H00695).

†ETH Zurich, 8092 Zurich, Switzerland, and University of Zurich, 8050 Zurich, Switzerland. e-mail: heinrich.nax@uzh.ch; website: nax.science.

‡Institute of Economic Research, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan. e-mail: newton@kier.kyoto-u.ac.jp; website: jonathannewton.net.

Level of rationality	Risk dominance with two strategies and uniform interaction	Generalized risk dominance
$k = 1$	Blume (2003)	Peski (2010)
$2 \leq k < \infty$	$*$	$*$
$k = \infty$	Myatt and Wallace (2003)	$*$

Figure 1: **Summary of literature.** Blume (2003) and Peski (2010) consider myopic ($k = 1$) players that follow the general class of processes that we consider in Section 6. Myatt and Wallace (2003) consider sophisticated ($k = \infty$) players who can iteratively reason to a Nash equilibrium every period. The current paper (denoted $*$) extends these results to any k in the generalized setting and, a fortiori, to the two strategy uniform interaction setting. In addition, our results span the table in the sense that if, as empirical work suggests (Stahl and Wilson, 1994; Nagel, 1995, onwards) and is theoretically plausible (Stahl, 1993), players have different k, or if levels of k are determined randomly from period to period, then the result still holds. Other work, discussed later, considers two player games under sample-based processes (Sáez-Martı and Weibull, 1999; Matros, 2003; Khan and Peeters, 2014).

Weibull, 1999; Myatt and Wallace, 2003), the literature on long run outcomes in games played in populations usually abstracts from such considerations. An open question has been whether the best known result in this literature, the emergence of (generalized) risk dominant Nash equilibria (see Peski, 2010) under broad classes of best response dynamics, is robust to iterative reasoning. We answer this question in the affirmative. Even though short run behavior can be dramatically affected by such reasoning and convergence to Nash equilibrium may fail to occur in the short run, long run predictions are robust to all levels of reasoning by players. Moreover, these levels may be heterogeneous and may even be random, in which case they can be correlated across players.

Let us describe our model. Every period, each player in a game is independently active with positive probability. Given the current strategy profile, an active player formulates a conjecture about the behavior of the other players to which he will usually, but not always, best respond. A player of level $k = 1$ will conjecture that other players remain at the current strategy profile. Higher levels of k are defined iteratively. A player of level k will conjecture that all other players are of level $k-1$. Level $k = \infty$ involves reasoning to a Nash equilibrium strategy every period. There are at least two ways to interpret such processes. One way is to take the model at face value. This is consistent with the foundational level k literature that explicitly considers behavioral dynamics (Nagel, 1995; Selten, 1991; Stahl, 1993). Another interpretation is that the game is played by new players each period, with these players looking to the actions of previous players for guidance. Some players may consider it sensible to best respond to the previous period's play, other players may expect their opponents to best respond to the previous period's play, and so on. This interpretation is consistent with level k thinking as a theory of behavior in one-off interactions that take place against a background of social learning. Note that under either interpretation, a level 1 player will conjecture that other players play as in the previous period. This contrasts with static models of level k thinking in which a level 1 player assumes that other players are 'level 0' players who choose strategies randomly and independently of how the game has been played in the past.

The strongest known results on the long run emergence of risk dominance under level k reasoning are shown in Figure 1, with the remaining entries in the table being contributions of the current paper. Given that risk dominance is robust across all entries in the table, it is worth explaining why the robustness of long run behavior to level k reasoning is not obvious. The reason is that short run behavior under level k dynamics can be very sensitive to differing levels of k. Short run behavior refers to behavior that will be observed with high probability when non-best responses are rare. Specifically, short run dynamics are the dynamics that we would observe if active players were to always best respond to their conjectures. Consider a coordination game between Alice and Bob in which each player has two strategies, A and B, and would like to choose the same strategy as the other player. Let both players have the same level of k. For $k = 1$, from any miscoordinated strategy profile, there is always a chance that one player is inactive while the other player best responds to the first player's current strategy. When this happens, coordination is attained. In contrast, for $k = 2$, miscoordination will persist. To see this, consider the profile where Alice plays A and Bob plays B. From this profile, a level 2 Alice will conjecture that Bob is level 1 and will best respond to Alice's current strategy by playing A. Therefore, Alice will best respond to her conjecture by continuing to play A. When Bob updates his strategy, he will reason similarly and will continue to play B. Persistent miscoordination is driven by persistently incorrect conjectures.[1]

Now, consider the influence of non-best response behavior over the long run. Specifically, consider the logit dynamics, under which strategies that lead to higher conjectured payoffs are more likely to be played. In contrast to the sensitivity of short run behavior, we find that long run stability of risk dominant equilibria under the logit dynamic is robust to different levels of k (Theorem 1). The intuition for this can be explained with reference to the above example. If A is the risk dominant strategy, then the probability of Alice playing A when she conjectures that Bob will play B is greater than the probability that she plays B when

[1] Note that this differs from the phenomenon of miscoordination due to synchronous strategic updating. For example, if Alice and Bob were both level 1 and always adjusted their strategies simultaneously, then from a starting position of miscoordination, their attempts to coordinate would repeatedly fail. In contrast, players' updating probabilities in our model are independent, so synchronicity is not a cause of miscoordination.

she conjectures that Bob will play A, regardless of whether her conjectures are correct or incorrect. That is, this bias exists regardless of whether short run behavior converges to a Nash equilibrium or involves persistent miscoordination. Consequently, long run behavior is pushed towards risk dominance, irrespective of the players' levels of rationality.

For expositional reasons, we first state our main result for a level k version of logit choice (Theorem 1). Applying Theorem 1 to technology adoption in networked populations, we show that the strategy profile at which every player plays a risk dominant technology remains the most stable outcome for arbitrary levels of rationality (Proposition 1). Later, we expand our analysis beyond the logit dynamic. Specifically, we consider a broad class of perturbed best response dynamics in which the probability of playing a non-best response is weakly decreasing in the payoff loss that results. In addition, we move beyond level k conjectures to consider a broad class of conjectures that players can make about the behavior of other players. Lastly, we consider randomness in players' conjectures (Theorem 2).

To prove these results, we use recent advances in the study of asymmetric dynamics. First, we break down the dynamic process of strategy updating into sub-processes for each player and level of rationality. For example, if Alice sometimes behaves as a level 1 player and sometimes as a level 2 player, then one of the sub-processes has Alice acting as a level 2 player while Bob's strategy is frozen and unchanging. Next, we show that these sub-processes satisfy a certain type of asymmetry towards generalized risk dominant strategy profiles. This asymmetry corresponds to the probabilistic bias in Alice's choices that we identified above. Importantly, the asymmetry is independent of the correctness or otherwise of the underlying conjectures. We then combine these sub-processes using the methods of Newton (2020, 2019) to obtain an aggregate process that is also asymmetric towards such profiles. This asymmetry implies that generalized risk dominant strategy profiles are those that will be observed most often in the long run (Peski, 2010).

The paper is organized as follows. Section 2 discusses related literature. Section 3 describes level k logit dynamics and Section 4 gives the first iteration of our main result. Section 5 applies this result to technology adoption on networks. Section 6 gives our main result under a broad class of dynamics and a broad class of conjectures that players can make about the behavior of other players. Section 7 concludes. All proofs are relegated to the appendix.

2. Related literature

Economic experiments on static, sequential and repeated games have produced evidence to support both level k modeling assumptions and the possibility of heterogeneity in k across players. See Crawford (2019) for a recent review of this literature. Moreover, there is evidence that short run convergence to equilibrium may depend on cognitive ability (Gill and Prowse, 2016; Proto et al., 2019). However, there is also experimental evidence of a tendency for subjects to reason at progressively higher levels over time (Nagel, 1995; Duffy and Nagel, 1997). Indeed, changes in levels of k have been fitted to models of reinforcement learning Stahl (1999, 2000) and Bayesian updating (Ho and Su, 2013). In contrast, the main results of the current paper concern long run behavior after levels of k have either reached a steady state or have risen to and persist above a given level. In such situations, our results predict a tendency towards risk dominant equilibria, as has been observed in several experimental settings (Van Huyck et al., 1990; Battalio et al., 2001; Heinemann et al., 2004; Cabrales et al., 2007).

There also exists a theoretical literature on the persistence of different levels of k in populations. The overall conclusion is that because “being right is just as good as being smart” (Stahl, 1993), heterogeneous levels of rationality can persist when players with those levels play similar strategies (see also Mohlin, 2012). Other work discusses explicit weaknesses of iterated reasoning. Stennek (2000) shows how iterated deletion of strictly dominated strategies can lead to fitness losses unless probability weight that is redistributed from a dominated strategy when it is deleted is only transferred to those strategies that dominate it. Geanakoplos and Gray (1991) explain how errors in assessing the value of future continuation games can lead to suboptimal play in the present. A striking way to be unable to assess the value of the future in such settings would be if a player did not use information about other player’s payoffs when making decisions. As we might expect, such players will usually be eliminated from the population (Robalino and Robson, 2016). Finally, heterogeneity in other traits may interact with iterated reasoning. For example, Heller (2015) shows how being able to know far in advance when a series of repeated prisoner’s dilemmas will end can be evolutionarily selected against, as when two such players are paired, it becomes impossible to sustain cooperation for a considerable number of periods before the end of the game. Similarly, collaborative decision making may lead to cooperation in prisoner’s dilemmas in the absence of further reasoning, but will fail to do so if subsequent reasoning leads to defection (Newton, 2017; Rusch, 2019).

The dynamic processes that we adapt for level k and broader conjectures about opponents’ behavior are common in the evolutionary literature. Theorem 1 concerns logit dynamics (Blume, 1993). For a detailed discussion of logit dynamics, the reader is directed to Alós-Ferrer and Netzer (2010). The general idea is that the probability of playing a non-best response strategy decreases log-linearly in the payoff loss from playing that strategy relative to the payoff from playing a best response. If log-linearity is dropped, we obtain a much larger class of dynamics. Theorem 2 concerns this class, which is close to the classes of *skew-symmetric* rules (Blume, 2003) and *payoff-based* rules (Peski, 2010). Recent experimental studies designed to explicitly test the properties of non-best response behavior find evidence in support of such dynamics (Mäs and Nax, 2016; Lim and Neary, 2016; Hwang et al., 2018). The survey of Newton (2018) covers recent work on trait evolution and dynamics, including many of the papers discussed above.

The dynamics considered in the current paper have the current strategy profile as the state variable. This has been popular in the literature following Kandori et al. (1993). A parallel literature considers a sample-based process, adaptive play (Young, 1993a), according to which members of populations are drawn to play a game against members of other populations and respond to a sample of how the game has been played in the recent past. This adds a degree of complication to the model that has been leveraged to obtain results for level k thinking in two player games in which one player is drawn from each of two populations. From a benchmark in which every player has $k = 1$, the focus of research has been on conditions under which the presence of $k = 2$ players changes the implications of the benchmark model. Sáez-Martı and Weibull (1999) consider the bargaining model of Young (1993b) in the presence of $k = 2$ players, Matros (2003) considers generic two player games in the presence of $k = 2$ players, and Khan and Peeters (2014) consider generic two player games in the presence of players with any finite k. The general conclusion of this literature is that $k > 1$ makes a difference if and only if a low sample size for clever players in one population causes them to foresee a change in the behavior of the opposing population the next period that does not in fact happen. However, having acted to preempt the predicted change, the clever players put in motion a sequence of transitions that moves the process to another equilibrium.

Finally, we note that there is a literature that considers possible non-convergence of certain processes to mixed Nash equilibria (see Crawford, 1974) and the role that $k = 2$ players can have in overcoming this non-convergence (Selten, 1991; Conlisk, 1993b,a; Tang, 2001). In contrast, the dynamics of the current paper may fail to converge to Nash equilibrium in the short run due to the presence of $k = 2$ players. However, our main results regarding long run predictions turn out to be unaffected by whether or not short run convergence occurs, thus providing a unifying theme to the literature described in Figure 1.

3. Model

3.1 The game

Consider a normal form game $G = (N, \{S_i\}_{i\in N}, \{U_i\}_{i\in N})$. The set of players is N. Each player $i \in N$ has a finite strategy set S_i and a strategy for player i is denoted $s_i \in S_i$. The set of strategy profiles is $S = \times_{i\in N} S_i$ with generic element $s \in S$. Let s_{-i} denote s restricted to $N \setminus \{i\}$. The payoff to player i at strategy profile s is given by $U_i(s) = U_i(s_i, s_{-i})$. Assume that $U_i(s_i, s_{-i}) \neq U_i(s'_i, s_{-i})$ for any s_i, s'_i, s_{-i}, which holds generically in payoffs. Cases in which it does not hold will be considered later in Section 6.

3.2 Level k best responses

We assume that each player $i \in N$ has a level of rationality given by an integer $k_i \geq 1$. Different players are allowed to have different values of k_i. We refer to a player with a given value of k_i as a *level k_i player*. A player's level will determine the conjecture he makes about the behavior of other players. In Section 6 we will allow players' levels to be random, correlated and changing over time. For now, we assume that any given player's level remains fixed and unchanging.

For a given strategy profile s, we denote the profile of best responses by

(1) $$B^1(s) = \Big(B_i^1(s)\Big)_{i\in N}, \quad \text{where} \quad B_i^1(s) \in \arg\max_{s_i \in S_i} U_i\Big(s_i, s_{-i}\Big).$$

We will refer to $B^1(s)$ as the profile of *level* 1 best responses to s. Note that, by our genericity assumption on payoffs, best responses are uniquely determined. We also wish to consider best responses to best responses, best responses to these in turn, and so on. To do this, we recursively define *level k* best responses as

(2) $$B^k(s) = B^1\Big(B^{k-1}(s)\Big) \quad \text{for} \quad k \in \mathbb{N},\ k \geq 2.$$

Note that the difference between different levels of best response lies in the conjectured strategy profiles to which a player best responds. The conjectures described here are based on the iteration of the best response correspondence.[2] More general conjectures that are not based on iterated best response will be considered in Section 6.

If the best response correspondence converges, that is if there exists k such that $B^k(s) =$

[2]Note that level 1 best responses are a best response to the status quo. Therefore, level 0 behavior in this context corresponds to *inertia*, in contrast to static models in which level 0 is often modeled as uniform random choice.

$$i \text{ is } \begin{cases} \text{active w.p. } q_i & \text{and plays } s_i' \text{ w.p. } \begin{cases} Pr^1(s_i'|s^t) & \text{if } i \text{ is level } 1 \\ Pr^k(s_i'|s^t) & \text{if } i \text{ is level } k_i \geq 2 \end{cases} \\ \text{inactive w.p. } (1-q_i) & \text{and plays } s_i^{t+1} = s_i^t \end{cases}$$

$$\xrightarrow[\text{Period } t]{} \quad \xrightarrow[\text{Period } t+1]{\hspace{20em}} \quad \xrightarrow[\text{Period } t+2]{}$$

Figure 2: **Strategy updating.** Revision probabilities in period $t+1$ for different levels of rationality of player i.

$B^{k-1}(s)$ for all $s \in S$, then we say that the game G is *Nash convergent*. In such games, iterative reasoning based on best responses will lead players to a Nash equilibrium.

Remark 1. If G is Nash convergent, then for all $s \in S$, $B^{\infty}(s) := \lim_{k\to\infty} B^k(s)$ is a Nash equilibrium (Nash, 1950).

3.3 Level k logit choice

For our principal analysis we will adopt the logit choice rule. We do this as it is the best known and most commonly used stochastic behavioral rule. However, our results extend to a family of payoff dependent behavioral rules, which we shall discuss in Section 6. The standard logit choice rule (Cox, 1958; Block and Marschak, 1959) is as follows. Given a current strategy profile s^t, the probability of player i choosing strategy s_i' is given by

$$Pr^1(s_i'|s^t) = \frac{e^{\frac{1}{\eta}U_i\left(s_i', s_{-i}^t\right)}}{\sum_{s_i \in S_i} e^{\frac{1}{\eta}U_i\left(s_i, s_{-i}^t\right)}}, \quad \text{for some } \eta > 0. \tag{3}$$

This is a perturbed best response rule parameterized by $\eta > 0$. The probability of choosing any given strategy under this rule is increasing in the payoff from choosing that strategy. For small values of η, a player following this rule will usually play a level 1 best response. Hence, we refer to the rule as *level* 1 *logit choice*. However, sometimes the player will play a non-best response. For small values of η, such non-best responses are rare and the level 1 best response is played almost all of the time. Analogously, we define *level k logit choice* for $k \geq 2$,

$$Pr^k(s_i'|s^t) = \frac{e^{\frac{1}{\eta}U_i\left(s_i', B_{-i}^{k-1}(s^t)\right)}}{\sum_{s_i \in S_i} e^{\frac{1}{\eta}U_i\left(s_i, B_{-i}^{k-1}(s^t)\right)}}. \tag{4}$$

That is, for small values of η, a player following the level k logit choice rule will usually play a level k best response.

For $\eta = 0$, define level k logit choice probabilities as the limits of (3) and (4) as $\eta \to 0$. That is, a level k best response will be played with probability one.

The difference between the standard logit choice rule and the level k logit choice rule for $k \geq 2$ is the conjectured play of the opposing players. Specifically, standard logit probabilities for player i are calculated with respect to the conjecture that other players remain playing their current strategies, whereas level k logit probabilities are calculated with respect to the conjecture that other players play level $k-1$ best responses.

It follows from (4) and the definition of Nash convergence that if the game G is Nash convergent and players are sufficiently rational, then level k logit choice under small η will usually conform to the play of Nash equilibrium strategies.[3]

Remark 2. If the game G is Nash convergent then, for large enough k_i, for any current strategy profile $s^t \in S$, level k_i logit choice by player i will choose the Nash equilibrium strategy $B_i^\infty(s^t)$ with probability approaching one as $\eta \to 0$.

3.4 Dynamic strategy updating

We define the *level k logit dynamics* on the state space of strategy profiles. The game is played repeatedly in discrete time and strategies are updated according to the level k logit choice rule. Let the strategy profile played at period t be s^t. At time $t+1$, any given player i is, independently of the other players, *active* with probability $q_i \in (0,1)$. If i is not active at $t+1$, then his strategy at $t+1$ remains the same as his strategy at t. That is, $s_i^{t+1} = s_i^t$. If player i is active at period $t+1$ and is of level k_i, then he updates his strategy according to the level k_i logit choice rule.

Level k logit dynamics combine the commonly used logit dynamic and level k conjectures. As remarked in the Introduction, there are two ways of interpreting this fusion. One way is to interpret the dynamic as a fixed set of players that reason in an iterative manner and continue to do so as time passes. Another interpretation is to consider players as being drawn from a population every period, observing the current strategy profile, using level k reasoning to form a conjecture about how other players will play, responding to this in a single iteration of the game, then disappearing back into the population to be replaced by somebody else the following period. This latter interpretation is consistent with level k thinking as a theory of behavior in one-off interactions, together with a process of social learning in which a society converges, or perhaps fails to converge, to a convention.

[3]In fact, it can be checked that, given current strategy profile s^t, logit choice probabilities under a sequence of decreasing values of η correspond to a sequence that identifies $B^\infty(s^t)$ as a proper equilibrium under the definition of Myerson (1978).

Remark 3. If $k_i = 1$ for every player, then this process is effectively the standard logit dynamic of Blume (1993).[4] Players for whom $k = 2$ correspond to the 'clever' players of Sáez-Martı and Weibull (1999); Matros (2003).

Remark 4. Let G be Nash convergent and every player have a level k_i high enough that $B^{k_i-1}(\cdot) = B^{\infty}(\cdot)$. Under the unperturbed ($\eta = 0$) dynamic, if the strategy profile at time t is s^t, then with probability at least $\prod_{i\in N} q_i$, the strategy profile at time $t+1$ will be the Nash equilibrium $B^{\infty}(s^t)$. High rationality players that achieve such coordination correspond to the 'sophisticated' players of Myatt and Wallace (2003).

Remarks 3 and 4 illustrate that the level k logit dynamics bridge the gap between perturbed best response dynamics in the style of Kandori et al. (1993); Young (1993a) and instantaneously jumping to a Nash equilibrium. Indeed, an important implication of the current paper is that certain results are robust across this entire class of models.

3.5 Stochastic stability

Under the level k logit dynamics with $\eta > 0$, any state can be reached from any other state. Therefore the process is irreducible and has a unique stationary distribution, which we denote π_η. The stationary distribution gives the proportion of time that the process will spend at any strategy profile in the long run. We are interested in dynamics that are close to best response dynamics, that is when η is small.

It can be shown by standard arguments (Foster and Young, 1990) that π_η converges as $\eta \to 0$. Denote this limiting stationary distribution by π_0. If $\pi_0(s) > 0$, we say that s is *stochastically stable*. If $\pi_0(s) = 1$, we say that s is *uniquely stochastically stable*. For small values of η, the process will spend almost all of its time at stochastically stable strategy profiles. Thus the identity of stochastically stable states tells us which strategy profiles we can expect to be played most of the time in the long run.

Regularities in behavior that are not observable in the short run may emerge in the long run. It is these regularities that we seek when we analyze long run behavior under perturbed dynamics. In the example in the introduction, unperturbed level 2 best response leads to every strategy profile being an absorbing state of the dynamics. In contrast, the drift towards A under the perturbed dynamic leads to the predominance of profile (A,A), as we shall see formally in the next section.

[4] See Alós-Ferrer and Netzer (2010) for an extended discussion of this process and Sandholm (2010); Newton (2018) for discussion of related processes.

	s^A	s	s'	s^B
Alice	A	A	B	B
Bob	A	B	A	B

Figure 3: A**-association.** Consider the illustrated example with two players and two strategies, labeled A and B, for each player. It can be seen that s^A is A-associated with every other strategy profile. Furthermore, s is A-associated with s', as every player plays A at at least one of these two profiles. However, s is not A-associated with s^B, as Bob plays A at neither of these profiles. Finally, s' and s^B are not A-associated, as Alice plays A at neither of these profiles.

4. Main result

Our main result is that results on the stability of risk dominant strategy profiles under standard ($k=1$) perturbed best response dynamics are robust to level k updating. We shall withhold additional discussion of why the result is novel, interesting and non-obvious until after presenting it. First, we shall define the concept of risk dominance that we use, *generalized risk dominance* (Peski, 2010).

Consider any given strategy profile and label it s^A. Without loss of generality, we label the strategies of every player at s^A as A, so that $s_i^A = A$ for all $i \in N$. If a pair of strategy profiles s, s' are such that every player plays A at at least one of s and s', then we say that s and s' are *A-associated* (see Figure 3). Generalized risk dominance of s^A holds when, for any two A-associated strategy profiles, the incentives to play A at one of these profiles outweighs any incentive not to play A at the other profile.[5]

Definition 1 (Generalized risk dominance)**.**
Profile s^A is *generalized risk dominant* (GR-dominant) if, for all A-associated strategy profiles s', s'', for all $i \in N$,

$$(5) \qquad U_i(A, s'_{-i}) + U_i(A, s''_{-i}) \geq \max_{s_i \neq A} U_i(s_i, s'_{-i}) + \max_{s_i \neq A} U_i(s_i, s''_{-i}).$$

Substituting $>$ for $\geq$ in (5) gives the definition of *strict generalized risk dominance*.

Remark 5. With two players and two strategies, (strict) GR-dominance is equivalent to (strict) risk dominance of Harsanyi and Selten (1988). We consider this further in Section 5. For two players and more than two strategies, (strict) GR-dominance implies (strict) $\frac{1}{2}$-dominance of Morris et al. (1995). Conversely, when there are many players but payoffs are

[5] Peski (2010) defines ordinal and cardinal forms of GR-dominance. The definition we use corresponds to the cardinal form.

a linear sum of payoffs from two player interactions, a strong form of $\frac{1}{2}$-dominance implies GR-dominance (Peski, 2010). In general, however, these concepts are independent (see also Iijima, 2015).

We are now ready to state our main Theorem. Risk dominance is robustly selected for under the entire class of level k logit dynamics.

Theorem 1. *Under the level k logit dynamics with $k_i \geq 1$ for all $i \in N$,*

- *If s^A is GR-dominant, then s^A is stochastically stable.*
- *If s^A is strictly GR-dominant, then s^A is uniquely stochastically stable.*

Remark 6. It is known that stochastic stability of GR-dominant profiles holds under a class of payoff dependent behavioral rules (see Section 6) that includes logit (Peski, 2010). This is a generalization of earlier results concerning the stochastic stability of risk dominant profiles in 2 by 2 games (Kandori et al., 1993; Young, 1993a; Blume, 1993, 2003). It is further known that these results are robust to heterogeneity in behavioral rules (Newton, 2020). However, unlike the above work, the current work considers non-myopic conjectures. Specifically, the above papers consider smoothed best responses to the current strategy profile s^t, whereas we consider smoothed best responses to a variety of conjectures based on s^t. Furthermore, we allow heterogeneity in these conjectures across players.

Remark 7. The seminal papers of Kandori et al. (1993); Young (1993a); Blume (1993) can be considered to have made two principal contributions. (I) The unperturbed ($\eta = 0$) dynamic eventually converges to a Nash equilibrium with probability one (under conditions of what Young calls *weak acyclicity*), and (II) in two strategy coordination games, risk dominant Nash equilibria are stochastically stable. In the current model, (I) does not hold. Persistent miscoordination may arise due to $k > 1$ and can prevent convergence of the unperturbed dynamic to a Nash equilibrium (see example in the Introduction and its formalization in Section 5). Nevertheless, result (II) continues to hold. That is, the long run stability of risk dominance does not rely on short run convergence to Nash equilibrium.

Remark 8. Consider Nash convergent G, small η, q_i close to 1 and k_i high enough that $B^{k_i-1}(\cdot) = B^{\infty}(\cdot)$ for all i. Under these conditions, from any non-Nash equilibrium profile, the process will reach a Nash equilibrium in a single period with a probability close to 1. This instant convergence, often assumed in one-shot games, does not change the stochastic stability of risk dominance. Taken together with Remark 7, this implies that stochastic stability of risk dominance is unaffected by either of the contrasting cases of instant convergence or non-convergence of the unperturbed dynamic to Nash equilibrium.

Remark 9. We show in Section 6 that our results are robust to the levels of players being generated randomly in each period in a way that allows for correlation, both positive and negative, across players. Consider an alternative approach of adding a state variable that tracks players' levels, with levels increasing over time. Considering this state variable as part of the state space, the process is no longer irreducible. However, if G is Nash convergent, then the process governing evolution of the strategy profile converges as best responses converge to $B^{\infty}(\cdot)$. This implies that the behavioral implications of Theorem 1 continue to hold.

Intuition behind the Proof. The proof of Theorem 1 is given in Appendix A. A summary of the proof is as follows. First, we disaggregate the process and consider sub-processes in which only a single player of some given level k updates his strategy, with the strategies of other players remaining fixed. Note that, unlike the aggregate process, such sub-processes are not irreducible. However, we can still show that they satisfy a particular property. Specifically, we show in Lemma 2 that if s^A is GR-dominant, then these sub-processes satisfy a form of asymmetry towards strategy A. This form of asymmetry was considered for aggregate processes by Peski (2010) and applied to individual behavioral rules by Newton (2020).

In the context of Theorem 1, asymmetry towards risk dominant profiles arises from two biases. Firstly, the logit dynamics have a probabilistic bias towards risk dominant profiles. Secondly, level k conjectures exhibit bias towards risk dominant profiles in the following sense: if profiles s' and s'' are A-associated and s^A is GR-dominant, then the conjectures $B_i^{k-1}(s')$ and $B_i^{k-1}(s'')$ that a level k player makes when faced with these strategy profiles are themselves A-associated. Importantly, this bias is present for any level k conjecture, regardless of its accuracy. This fact is what makes Theorem 1 independent of players' levels.

The next step in the proof is to re-aggregate the sub-processes to once again consider the aggregate process in which every player updates independently as described in our model. We show in Lemma 3 that asymmetry of the disaggregated processes implies asymmetry of the aggregate process. This is done by applying Theorem 3 of Newton (2020, 2019). Finally, we apply Theorem 1 of Peski (2010), which states that asymmetry of the aggregate process towards strategy A implies stochastic stability of s^A.

5. Application: technology adoption on networks

Consider a situation in which each player may adopt one of two technologies. Specifically, let $S_i = \{A, B\}$ for all $i \in N$. Each player is influenced by other players and may be influenced by some players more than others. Let $w_{ij} \in \mathbb{R}_{\geq 0}$ be the influence of player j on player i. Assume that every player i is influenced by at least one other player, so that

	A	B
A	a_{AA}	a_{AB}
B	a_{BA}	a_{BB}

Figure 4: **Unweighted payoffs from each interaction.** $a_{AA} > a_{BA}$ and $a_{BB} > a_{AB}$. For each combination of A and B, entries give payoffs for the row player.

$\sum_{j \in N \setminus \{i\}} w_{ij} > 0$. Each player wishes to adopt a similar strategy to those who influence him. Specifically, payoffs from each pairwise interaction are given by the game illustrated in Figure 4. The payoff to player i at strategy profile s is then the sum of the payoffs from his pairwise interactions weighted by their influence. That is,

$$(6) \qquad U_i(s) = U_i(s_i, s_{-i}) = \sum_{j \in N \setminus \{i\}} w_{ij} a_{s_i s_j},$$

where $a_{s_i s_j} \in \mathbb{R}$ is the payoff to player i from his interaction with player j.

A classic result (Blume, 1993) is the stochastic stability of strategy profiles at which every player plays the same risk dominant strategy (Harsanyi and Selten, 1988). In the two player game of Figure 4, a strategy is risk dominant if it maximizes payoff given that the opposing player randomizes evenly across his two strategies.

Definition 2.
Strategy A is *risk dominant* if

$$(7) \qquad a_{AA} + a_{AB} \geq a_{BA} + a_{BB},$$

and is *strictly risk dominant* if the inequality holds strictly.

It is possible to show that (strict) risk dominance of A implies (strict) GR-dominance of s^A. We can then apply Theorem 1 to show that (strict) risk dominance of A implies (unique) stochastic stability of s^A. The reverse implication then follows from the fact that at least one of the two strategies must be risk dominant.

Proposition 1. *For technology adoption on a network under the level k logit dynamics with* $k_i \geq 1$ *for all* $i \in N$

- s^A *is stochastically stable if and only if A is risk dominant.*
- s^A *is uniquely stochastically stable if and only if A is strictly risk dominant.*

The special case of Proposition 1 in which all players are level 1 is known from Blume (1993, 1996). Proposition 1 shows that stochastic stability of risk dominance is robust to varying levels of rationality. The result is not obvious. We saw in the Introduction that level k thinking can lead to short run behavior that is completely different to that implied by level 1 thinking. However, in all cases, long run behavior tends to risk dominance (see also Remark 7). The following formalizes the example from the Introduction.

Example 1. Let $N = \{i, j\}$ and $w_{ij} = w_{ji} = 1$. If player i is of level 2, then he will never change his strategy as a result of playing a best response to his conjecture. To see this, let the current strategy profile be $s^t = (s_i^t, s_j^t)$. Given this current strategy profile, player i will conjecture that player j will play $B_j^1(s^t) = s_i^t$. That is, player i expects player j at time $t+1$ to coordinate with the strategy of player i at time t. A best response for player i to this conjecture is to remain playing the same strategy at time $t+1$ as he plays at time t. That is, he does not change his strategy. Consequently, if both player i and player j are level 2, then neither of them will ever change his strategy as a result of a best response. It follows that all strategy profiles are absorbing states of the process with $\eta = 0$. This is in stark contrast to the standard case in which every player has level 1, where the process with $\eta = 0$ converges with probability one in finite time to a Nash equilibrium of the game. Nevertheless, Proposition 1 tells us that level k does not affect stochastic stability predictions for the perturbed process. The stability of risk dominance is robust to rationality.

6. Generalization

In this section we generalize the model in several dimensions. As before, we consider players who do not best respond to the current strategy profile, but rather form conjectures about play at $t+1$ to which they best respond. However, now we do not restrict attention to level k conjectures, but rather consider a more general class. In general, a *conjecture* for player i is a function $f_i : S \to S$. A profile of conjectures is given by $f = \{f_i\}_{i \in N}$.

An important class of conjectures are those that preserve A-association. Given A-associated profiles s, s', this requires that the respective conjectures formed when presented with these strategy profiles are themselves A-associated.

Definition 3 (A-association preserving)**.** Profile of conjectures f *preserves A-association* if, for all $i \in N$, s, s' A-associated, we have that $f_i(s), f_i(s')$ are A-associated.

The conjectures considered so far in the paper correspond to $f_i(s) = s$ and $f_i(s) = B^{k-1}(s)$ for $k \geq 2$. The conjecture $f_i(s) = s$ always satisfies Definition 3 (see below). For $f_i(s) = B^{k-1}(s)$, if s^A is GR-dominant and best responses are unique, then Definition 3 is satisfied

(Lemma 1 in Appendix A). The intuition is that, when s^A is GR-dominant, expression (5) implies that for A-associated s', s'', for all $i \in N$, strategy A is a (unique) best response to at least one of s' or s''. Therefore, for all $i \in N$, we have $B_i^1(s') = A$ or $B_i^1(s'') = A$, so that $B^1(s')$ and $B^1(s'')$ are A-associated. This logic continues as we iterate the best response operator. Given that this was our only use of the assumption of unique best responses, for the remainder of the paper we drop this assumption.

We shall show in Theorem 2 that our main result continues to hold when we replace level k conjectures with those satisfying Definition 3. However, the emphasis we have placed on level k conjectures has not been arbitrary. Indeed, given that the dynamics we consider are built around best response to a conjecture, the most natural conjectures to consider are those that are themselves constructed using best response. Furthermore, a consequence of this shared reliance on best response is that GR-dominance suffices to bias both the conjectures (via Definition 3) and strategy choice in favour of the GR-dominant strategies.

We present some examples of conjectures that satisfy Definition 3 irrespective of whether GR-dominance holds.

Example 2 (Myopia)**.** Consider the conjecture that all players remain playing their current strategy. That is, $f_i(s) = s$. It follows trivially that if s, s' are A-associated, then $f_i(s), f_i(s')$ are A-associated.

Example 3 (Majoritarianism)**.** Let $|N|$ be odd and consider the conjecture that all players play the most popular current strategy, with some tie breaking rule employed. If s, s' are A-associated, it must be that a majority of players at s or s' play A. Consequently, $f_i(s) = s^A$ or $f_i(s') = s^A$, therefore $f_i(s), f_i(s')$ are A-associated.

Example 4 (Imitate a friend)**.** Consider the conjecture that each player imitates some other player. That is, for all $j \in N$, we have $(f_i(s))_j = s_{k_j}$ for some $k_j \in N$. If s, s' are A-associated, then $s_{k_j} = A$ or $s'_{k_j} = A$. Consequently, for all j, $(f_i(s))_j = A$ or $(f_i(s'))_j = A$. Therefore, $f_i(s), f_i(s')$ are A-associated.

We present some examples of conjectures that fail to satisfy Definition 3. The first is the polar opposite of Example 3, the second a conjecture based on averaging.

Example 5 (Minoritarianism)**.** Consider the conjecture that all players play the least popular current strategy, with some tie breaking rule employed. Consider s, s' that are A-associated, with a majority of players at both s and s' playing A. Consequently, for all $j \in N$, we have $(f_i(s))_j \neq A$ and $(f_i(s'))_j \neq A$, therefore $f_i(s), f_i(s')$ are not A-associated.

Example 6 (Averaging)**.** Let $|N|$ be even and, for all $i \in N$, let $S_i = \{0, 1, 2\}$. Let $A \equiv 0$. Consider the conjecture that all players play the average strategy, rounded to the nearest integer. That is, for all $j \in N$, $(f_i(s))_j = \lfloor \frac{1}{|N|} \sum_{i \in N} s_i + \frac{1}{2} \rfloor$. Consider s, s' that are A-associated, with $s_i = 0, s'_i = 2$ for $i = 1, \ldots, \frac{|N|}{2}$ and $s_i = 2, s'_i = 0$ for $i = \frac{|N|}{2} + 1, \ldots, |N|$. Then, for all $j \in N$, we have $(f_i(s))_j = (f_i(s'))_j = 1$, therefore $f_i(s), f_i(s')$ are not A-associated.

Another dimension that we generalize is the behavioral rule that acts on the conjectured strategy profile. So far, we have considered the logit choice rule. Now, we consider a large class of *regular* behavioral rules (Young, 1993a) in which behavior depends on payoff differences.[6]

Adjust the behavioral rule given in Section 3 as follows. If player i is active at period $t+1$, then he forms a conjecture $f_i(s^t)$ about behavior at time $t+1$. Given his conjecture about the behavior of other players, he then updates his strategy according to a perturbed best response rule parameterized by η. Let $\Upsilon_i(s'_i, s'')$ denote the expected payoff loss, relative to a best response, incurred by player i when he plays s'_i against s''_{-i}. That is,

$$\Upsilon_i(s'_i, s'') = \max_{s_i \in S_i} U_i\left(s_i, s''_{-i}\right) - U_i\left(s'_i, s''_{-i}\right) \tag{8}$$

The probability that $s_i^{t+1} = s'_i$ at time $t+1$ is then given by

$$Pr\left(s_i^{t+1} = s'_i | s^t\right) = \Big(a + o(1)\Big)\, e^{-\frac{1}{\eta} g_i\left(\Upsilon_i\left(s'_i, f_i(s^t)\right)\right)}, \tag{9}$$

where a is a strictly positive constant that can depend on s'_i, s^t, f_i, but not on η; $o(1)$ is a term that approaches zero as $\eta \to 0$; g_i is a nonnegative, weakly increasing function. Parameter values should satisfy that the sum of probabilities over all $s'_i \in S_i$ equals one.

Note that larger values of $g_i(\cdot)$ in (9) imply smaller probabilities. Together with (8), this implies that the probability of choosing s'_i decreases in the difference between the payoff from best responding to the conjectured profile $f_i(s^t)$ and the payoff from playing s'_i against $f_i(s^t)$. The logit choice rule corresponds to $g_i(x) = x$ for appropriate choice of a and $o(1)$.

We present one final generalization. It may be that player's conjectures vary from period to period. It may even be the case that players' conjectures are correlated with each other. For example, it may be that a player is more likely to exhibit low rationality behavior when other players are exhibiting low rationality behavior.

Let F be a set of profiles of conjectures and let φ be an exogenously given distribution over F. Adjust the model so that rather than there being a single fixed profile of conjectures, every period a profile of conjectures f is chosen according to φ.

We are now ready to present our generalization of Theorem 1. If conjectures preserve

[6] These behavioral rules roughly correspond to *skew-symmetric* rules (Blume, 2003), *payoff-based* rules (Peski, 2010) and *self regarding payoff-difference based* rules (Newton, 2020).

A association, then under the class of perturbed best response rules described by (9), GR-dominance implies stochastic stability. Thus, the results of the current paper extend to a wide range of conjectures and behavioral rules.

Theorem 2. *Let all $f \in F$ preserve A-association. Under the dynamics (9)*

- *If s^A is GR-dominant, then s^A is stochastically stable.*
- *If s^A is strictly GR-dominant and, for all $i \in N$, we have g_i strictly increasing and $f_i(s^A) = s^A$ for all $f \in F$, then s^A is uniquely stochastically stable.*

We end this section with an example, applying our theorems to another popular behavioral rule that is covered by our general model. This is the *best response with uniform deviations* rule in which all perturbations from best response occur with similar probability.

Definition 4. A behavioral rule is *best response with uniform deviations* if

$$g_i(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 & \text{if } x > 0 \end{cases}. \tag{10}$$

For appropriate choice of a and $o(1)$, best response with uniform deviations corresponds to the best response with mutations rule of Kandori et al. (1993). The important thing about such rules is that any non-best response occurs with a probability of order $\varepsilon := e^{-\frac{1}{\eta}}$.

Example 7. Consider the technology adoption game of Section 5 played under the best response with uniform deviations rule given in Definition 4. Players are level k players whose levels are randomly determined every period. Specifically, let the set of possible profiles of conjectures F be such that for all $f \in F$, $i \in N$, $s \in S$, either $f_i(s) = s$ or $f_i(s) = B^{k-1}(s)$ for $k \geq 2$. For example, F might contain a low rationality profile of conjectures in which every player is of level 1 or 2, as well as a high rationality profile of conjectures in which every player is of level 3 or 4. If A is risk dominant, then s^A is GR-dominant (Step 1 of the proof of Proposition 1 in Appendix A). Therefore, the first part of Theorem 2 applies so that risk dominance of A implies that s^A is stochastically stable.

Proposition 2. *Consider Example 7. If A is risk dominant, then s^A is stochastically stable.*

Note that under best response with uniform deviations, g_i is not strictly increasing, so the second part of Theorem 2 cannot be similarly applied. To see this, consider the two player network in Example 1 and let A be strictly risk dominant so that s^A is strictly GR-dominant. Let s and s' be such that $s_i \neq s'_i$ and $s_j \neq s'_j$. That is, the strategies of both players differ across

s and s'. Note that A (respectively, B) is a best response to a level k_i conjecture on s if and only if B (respectively, A) is a best response to the same conjecture on s'. Further note that, by Definition 4, all non-best responses are played with the same order of probability. From this, we see that in terms of transition probabilities there is no difference between strategies A and B. The profile at which both players play A is stochastically stable by Proposition 2, therefore the profile at which both players play B is also stochastically stable. This possible failure of uniqueness is briefly discussed in the next section.

7. Discussion

7.1 Uniqueness

Best response with uniform deviations (Definition 4) does not transfer all payoff information relevant to a transition to the transition probabilities themselves. As we saw in Example 7, this insensitivity can lead to non-uniqueness of stochastically stable states even when there exists a strictly GR-dominant strategy profile. However, uniqueness may obtain if other aspects of the model make the aggregate dynamic sensitive to payoff differences. For example, we can obtain uniqueness in the technology adoption game if the network is such that, allowing a "free" mutation of any player to B, we still have GR-dominance of s^A in the restricted problem with that player's action fixed. This is essentially the logic that underpins the condition for uniqueness ("strict ordinal GR-dominance") given by Peski (2010) for $k = 1$.

7.2 Noisy introspection

Note that in our model of level k logit choice, players' conjectures do not consider the possibility that other players' choices are perturbed. This is in contrast to, for example, the 'noisy introspection' model of Goeree and Holt (2004). In an environment of large perturbations, this could lead to differences in best responses. For example, Alice may be concerned about the possibility of negative consequences arising from a perturbation in Bob's choice. In contrast, the model of the current paper considers small perturbations. As these perturbations vanish ($\eta \to 0$), noise in conjectures will also vanish, so that the payoffs that enter Alice's choice function will converge to those used in expressions (3) and (4) of our model.

7.3 Comparison to adaptive play

Consider the model of adaptive play (Young, 1993a) in which members of two populations (labeled α and β) sample and respond to past behavior. Let the sample sizes used by members of each population be s_α and s_β respectively and let $s_\alpha \leq s_\beta$ without loss of generality. There is a difference in long run behavior between the cases (i) $s_\alpha < s_\beta$ and $k = 1$ for all members of population α, and (ii) all other cases (see Khan and Peeters, 2014). That is, long term differences in behavior that arise due to differences in sample size are not robust to level k thinking. In contrast, for models for which the current strategy profile is the state variable, we uncover no sensitivity with regard to long run behavior (Theorem 1).

7.4 Afterword

In this article, we showed that GR-dominant strategies emerge in the long run when behavioral rules similar to best response are played against a large class of conjectures about the play of other players. This class includes conjectures that are themselves formulated using best response, in particular the conjecture that other players play level k best responses. The results hold despite the fact that short run behavior in the dynamic can be sensitive to model details. Natural avenues for further research include the study of alternative learning models, as well as the collection and interpretation of empirical data in light of the theory.

Appendix

A. Proofs

A.1 Additional definitions and useful results

For parameter value η, strategy profiles $s, s' \in S$, let $P^\eta(s, s')$ denote the probability that $s^{t+1} = s'$ conditional on $s^t = s$.

Define a new Markov process on S, denoted P_i^η, by adjusting the original process by setting $q_j = 0$ for all $j \neq i$. Let $P_i^\eta(s, s')$ denote the probability that $s^{t+1} = s'$ conditional on $s^t = s$. Observe that, for all $\eta > 0$, $s, s' \in S$, we have

$$(11) \qquad P^\eta\left(s, s'\right) = \prod_{i \in N} P_i^\eta\left(s, (s'_i, s_{-i})\right).$$

Define cost functions

$$(12) \qquad c_i(s, s') := \begin{cases} \lim_{\eta \to 0} -\eta \log P_i^\eta(s, s') & \text{if } P_i^{\hat{\eta}}(s, s') > 0 \text{ for some } \hat{\eta} > 0, \\ \infty & \text{otherwise}. \end{cases}$$

and let $c(s,s')$ be the equivalent quantity after replacing P_i^η by P^η.

Simple algebra shows that, for the updating rule in our model, we have

$$(13) \qquad c_i(s,s') := \begin{cases} 0 & \text{if } s'=s, \\ \max_{x_i\in S_i} U_i\left(x_i, B_{-i}^{k-1}(s)\right) - U_i\left(s_i', B_{-i}^{k-1}(s)\right) & \text{if } s_i' \neq s_i,\, s_{-i}'=s_{-i}, \\ \infty & \text{otherwise .} \end{cases}$$

We adopt the convention that $\infty > \infty$.

We say that s'' A-dominates s' if $s_i''=A$ for all i such that $s_i'=A$.

Definition 5. $c(\cdot,\cdot)$ is *asymmetric* (towards A) if, for any $s,s',\tilde{s}$ such that $s,\tilde{s}$ are A-associated, there exists $\tilde{s}'$ such that

- $\tilde{s}'$ A-dominates $\tilde{s}$,
- $s',\tilde{s}'$ are A-associated, and
- $c(s,s') \geq c(\tilde{s},\tilde{s}')$.

Definition 6. $c(\cdot,\cdot)$ is *strictly asymmetric* (towards A) if

(a) for any $s \neq s^A$, $c(s^A,s) > 0$, and

(b) for any $s,s',\tilde{s}$ such that $s,\tilde{s}$ are A-associated, there exists $\tilde{s}'$ such that
   - $\tilde{s}'$ A-dominates $\tilde{s}$,
   - $s',\tilde{s}'$ are A-associated, and
   - either $c(\tilde{s},\tilde{s}')=0$ or $c(s,s') > c(\tilde{s},\tilde{s}')$.

Note that Definitions 5 and 6 can be applied to both $c(\cdot,\cdot)$ and to $c_i(\cdot,\cdot)$.

Theorem N (Newton, 2020, 2019)**.**
Let P^η, $\{P_i^\eta\}_{i\in N}$ satisfy (11). ***(i)*** *If $c_i(\cdot,\cdot)$ is asymmetric for all $i\in N$, then $c(\cdot,\cdot)$ is asymmetric (Newton, 2020, Theorem 3).* ***(ii)*** *If $c_i(\cdot,\cdot)$ is strictly asymmetric for all $i\in N$, then $c(\cdot,\cdot)$ is strictly asymmetric (Newton, 2019, Theorem 3).*

Theorem P (Peski, 2010, Theorem 1)**.**
(i) *If $c(\cdot,\cdot)$ is asymmetric, then s^A is stochastically stable.* ***(ii)*** *If $c(\cdot,\cdot)$ is strictly asymmetric, then s^A is uniquely stochastically stable.*

A.2 Proofs for Section 4

Lemma 1. *Let $s, \tilde{s}$ be A-associated. If s^A is GR-dominant, then $B^k(s), B^k(\tilde{s})$ are A-associated for all $k \geq 1$.*

Proof. If $s, \tilde{s}$ are A-associated, then by GR-dominance of s^A, in particular expression (5), we have

$$U_i(A, \tilde{s}_{-i}) - \max_{x_i \neq A} U_i(x_i, \tilde{s}_{-i}) \geq \max_{x_i \neq A} U_i(x_i, s_{-i}) - U_i(A, s_{-i}). \tag{14}$$

If $B_i^1(s) \neq A$, we have

$$\max_{s_i \neq A} U_i(s_i, s_{-i}) - U_i(A, s_{-i}) > 0, \tag{15}$$

so combining (14) and (15) we obtain

$$U_i(A, \tilde{s}_{-i}) - \max_{x_i \neq A} U_i(x_i, \tilde{s}_{-i}) > 0. \tag{16}$$

Therefore, $B_i^1(\tilde{s}) = A$. This holds for all i such that $B_i^1(s) \neq A$, therefore $B_i^1(s), B_i^1(\tilde{s})$ are A-associated.

Iterating the above argument, we obtain that $B_i^k(s), B_i^k(\tilde{s})$ are A-associated for $k = 2, 3, \ldots$. □

Lemma 2. ***(i)*** *If s^A is GR-dominant, then, for all $i \in N$, c_i is asymmetric towards A.* ***(ii)*** *If s^A is strictly GR-dominant, then, for all $i \in N$, c_i is strictly asymmetric towards A.*

Proof. Note that as s^A is A-associated with itself, GR-dominance of s^A and uniqueness of best responses implies that

$$B^k(s^A) = s^A \quad \text{for all } k \geq 0. \tag{17}$$

If $s \neq s^A$, then either $s_{-i} \neq s^A_{-i}$, in which case $c_i(s^A, s) = \infty$, or $s_{-i} = s^A_{-i}$, $s_i \neq A$, in which case

$$\begin{aligned} c_i(s^A, s) &\underbrace{=}_{\text{by (13)}} \max_{x_i \in S_i} U_i(x_i, B_{-i}^{k-1}(s^A)) - U_i(s_i, B_{-i}^{k-1}(s^A)) \\ &\underbrace{=}_{\text{by (17)}} \max_{x_i \in S_i} U_i(x_i, s^A) - U_i(s_i, s^A) \\ &= U_i(A, s^A) - U_i(s_i, s^A) \\ &> 0. \end{aligned} \tag{18}$$

Therefore, the condition in Definition 6a is satisfied.

Now consider $s, s', \tilde{s} \in S$ such that $s, \tilde{s}$ are A-associated. If s^A is GR-dominant, it follows from Lemma 1 that

$$(19) \qquad B^k(s), B^k(\tilde{s}) \quad \text{are } A\text{-associated for all } k \geq 1.$$

<u>Case 1</u>: $s = s'$ or $s'_i = A$ or $c_i(s,s') = \infty$ or $\tilde{s}_i = A$.

If $c_i(s,s') = \infty$, let $\tilde{s}' = s^A$. The conditions in Definitions 5 and 6b are satisfied.

If $c_i(s,s')$ is finite, let $\tilde{s}' = \tilde{s}$. (13) implies $c_i(\tilde{s},\tilde{s}') = 0$, therefore the conditions in Definitions 5 and 6b are satisfied.

<u>Case 2</u>: $s \neq s'$ and $s'_i \neq A$ and $c_i(s,s')$ is finite and $\tilde{s}_i \neq A$.

(13) together with finiteness of $c_i(s,s')$ implies $s_{-i} = s'_{-i}$. $s, \tilde{s}$ are A-associated, so $\tilde{s}_i \neq A$ implies that $s_i = A$. $s \neq s'$ then implies $s'_i \neq A$. Let $\tilde{s}'$ be such that $\tilde{s}'_{-i} = \tilde{s}_{-i}$, $\tilde{s}'_i = A$.

If $c_i(\tilde{s},\tilde{s}') = 0$, then the conditions in Definitions 5 and 6b are satisfied.

If $c_i(\tilde{s},\tilde{s}') > 0$, then

$$
\begin{aligned}
(20) \qquad c_i(s,s') &\underbrace{=}_{\text{by (13)}} \max_{x_i \in S_i} U_i(x_i, B^{k-1}_{-i}(s)) - U_i(s'_i, B^{k-1}_{-i}(s)) \\
&\geq U_i(A, B^{k-1}_{-i}(s)) - \max_{x_i \neq A} U_i(x_i, B^{k-1}_{-i}(s)) \\
&\underbrace{\geq}_{\substack{\text{by (19)} \\ \text{and GR-dominance}}} \max_{x_i \neq A} U_i(x_i, B^{k-1}_{-i}(\tilde{s})) - U_i(A, B^{k-1}_{-i}(\tilde{s})) \\
&\underbrace{=}_{\substack{\text{by (13)} \\ \text{and } c(\tilde{s},\tilde{s}')>0}} c_i(\tilde{s},\tilde{s}').
\end{aligned}
$$

That is, the condition in Definition 5 as satisfied. Further note that if s^A is strictly GR-dominant, then the weak inequality in (20) due to GR-dominance becomes a strict inequality, so that the condition in Definition 6b is satisfied. $\square$

Lemma 3. ***(i)*** *If s^A is GR-dominant, then c is asymmetric towards A.* ***(ii)*** *If s^A is strictly GR-dominant, then c is strictly asymmetric towards A.*

Proof. **(i)** GR-dominance of s^A and Lemma 2(i) together imply that, for all $i \in N$, c_i is asymmetric towards A. Theorem N(i) then implies that c is asymmetric towards A. **(ii)** Strict GR-dominance of s^A and Lemma 2(ii) together imply that, for all $i \in N$, c_i is strictly asymmetric towards A. Theorem N(ii) then implies that c is strictly asymmetric towards A. $\square$

Proof of Theorem 1.
Assume s^A is GR-dominant. Lemma 3(i) and Theorem P(i) together imply stochastic stability of s^A.

Assume s^A is strictly GR-dominant. Lemma 3(ii) and Theorem P(ii) together imply unique stochastic stability of s^A. □

A.3 Proofs for Section 5

Proof of Proposition 1.

Step 1

First we show that if A is (strictly) risk dominant, then s^A is (strictly) GR-dominant. Note that condition (5) for GR-dominance reduces to

$$U_i(A,s'_{-i})+U_i(A,s''_{-i}) \geq U_i(B,s'_{-i})+U_i(B,s''_{-i}) \tag{21}$$

for all $i \in N$, s', s'' A-associated.

Now, if s', s'' are A-associated, then for all i,

$$\begin{aligned}
&U_i(A,s'_{-i})-U_i(B,s'_{-i}) \\
&\underbrace{=}_{\text{by (6)}} \sum_{\substack{j\in N\setminus\{i\} \\ s'_j=A}} w_{ij}\,(a_{AA}-a_{BA}) - \sum_{\substack{j\in N\setminus\{i\} \\ s'_j=B}} w_{ij}\,(a_{BB}-a_{AB}) \\
&\underbrace{\geq}_{\text{by (7)}} \sum_{\substack{j\in N\setminus\{i\} \\ s'_j=A}} w_{ij}\,(a_{BB}-a_{AB}) - \sum_{\substack{j\in N\setminus\{i\} \\ s'_j=B}} w_{ij}\,(a_{AA}-a_{BA}) \\
&\underbrace{\geq}_{\substack{\text{by } A\text{-association} \\ \text{of } s', s''}} \sum_{\substack{j\in N\setminus\{i\} \\ s''_j=B}} w_{ij}\,(a_{BB}-a_{AB}) - \sum_{\substack{j\in N\setminus\{i\} \\ s''_j=A}} w_{ij}\,(a_{AA}-a_{BA}) \\
&\underbrace{=}_{\text{by (6)}} U_i(B,s''_{-i})-U_i(A,s''_{-i}).
\end{aligned} \tag{22}$$

That is, (21) holds and s^A is GR-dominant. If risk dominance is strict, then the first $\geq$ in (22) is strict, therefore (21) holds strictly and s^A is strictly GR-dominant.

Step 2

The definition of risk dominance implies that at least one of A, B is risk dominant.

Non-strict

By Step 1, if A is risk dominant, then s^A is GR-dominant. Theorem 1 then implies that s^A is stochastically stable.

If A is not risk dominant, then B is strictly risk dominant. Step 1 then implies that s^B is strictly GR-dominant and Theorem 1 implies that s^B is uniquely stochastically stable. Therefore, s^A is not stochastically stable.

Strict

By Step 1, if A is strictly risk dominant, then s^A is strictly GR-dominant. Theorem 1 then implies that s^A is uniquely stochastically stable.

If A is not strictly risk dominant, then B is risk dominant. Step 1 then implies that s^B is GR-dominant and Theorem 1 implies that s^B is stochastically stable. Therefore, s^A is not uniquely stochastically stable. □

A.4 Proofs for Section 6

Applying (12) to (9), we have

$$
(23) \qquad c_i(s,s') := \begin{cases} 0 & \text{if } s' = s, \\ g_i\left(\Upsilon_i(s'_i, f_i(s))\right) & \text{if } s'_i \neq s_i,\ s'_{-i} = s_{-i}, \\ \infty & \text{otherwise .} \end{cases}
$$

Lemma 4. *Let f preserve A-association. **(i)** If s^A is GR-dominant, then, for all $i \in N$, c_i is asymmetric towards A. **(ii)** If s^A is strictly GR-dominant and, for all $i \in N$, g_i is strictly increasing and $f_i(s^A) = s^A$, then, for all $i \in N$, c_i is strictly asymmetric towards A.*

Proof. Consider $s, s', \tilde{s} \in S$ such that $s, \tilde{s}$ are A-associated. As f preserves A-association, we have

$$
(24) \qquad f_i(s), f_i(\tilde{s}) \quad \text{are } A\text{-associated.}
$$

Case 1: $s = s'$ or $s'_i = A$ or $c_i(s,s') = \infty$ or $\tilde{s}_i = A$.

If $c_i(s,s') = \infty$, let $\tilde{s}' = s^A$. The conditions in Definitions 5 and 6b are satisfied.

If $c_i(s,s')$ is finite, let $\tilde{s}' = \tilde{s}$. (23) implies $c_i(\tilde{s},\tilde{s}') = 0$, therefore the conditions in Definitions 5 and 6b are satisfied.

Case 2: $s \neq s'$ and $s'_i \neq A$ and $c_i(s,s')$ is finite and $\tilde{s}_i \neq A$.

(23) and finiteness of $c_i(s,s')$ implies $s_{-i} = s'_{-i}$. $s, \tilde{s}$ are A-associated, so $\tilde{s}_i \neq A$ implies that $s_i = A$. $s \neq s'$ then implies $s'_i \neq A$. Let $\tilde{s}'$ be such that $\tilde{s}'_{-i} = \tilde{s}_{-i}$, $\tilde{s}'_i = A$.

If $c_i(\tilde{s},\tilde{s}') = 0$, then the conditions in Definitions 5 and 6b are satisfied.

If $c_i(\tilde{s},\tilde{s}') > 0$, then

$$
\begin{aligned}
(25)\qquad c_i(s,s') &\underbrace{=}_{\text{by (23)}} g_i\left(\Upsilon_i(s_i', f_i(s))\right)\\
&\underbrace{=}_{\text{by (8)}} g_i\left(\max_{x_i\in S_i} U_i(x_i,(f_i(s))_{-i}) - U_i(s_i',(f_i(s))_{-i})\right)\\
&\geq g_i\left(U_i(A,(f_i(s))_{-i}) - \max_{x_i\neq A} U_i(x_i,(f_i(s))_{-i})\right)\\
&\underbrace{\geq}_{\substack{\text{by (24)}\\ \text{and GR-dominance}}} g_i\left(\max_{x_i\neq A} U_i(x_i,(f_i(\tilde{s}))_{-i}) - U_i(A,(f_i(\tilde{s}))_{-i})\right)\\
&\underbrace{=}_{\text{by (8)}} g_i\left(\Upsilon_i(\tilde{s}_i', f_i(\tilde{s}))\right) \underbrace{=}_{\substack{\text{by (23)}\\ \text{and } c(\tilde{s},\tilde{s}')>0}} c(\tilde{s},\tilde{s}').
\end{aligned}
$$

That is, the condition in Definition 5 as satisfied. Further note that if s^A is strictly GR-dominant and g_i is strictly increasing, then the weak inequality in (25) due to GR-dominance becomes a strict inequality, so that the condition in Definition 6b is satisfied.

For the remainder of this proof, assume that $f(s^A) = s^A$, s^A is strictly GR-dominant, and g_i is strictly increasing.

Note that, as s^A is A-associated with itself, strict GR-dominance of s^A implies that

$$
(26)\qquad \arg\max_{x_i\in S_i} U_i(x_i,(s^A)_{-i}) = \{A\}.
$$

If $s \neq s^A$, then either $s_{-i} \neq s^A_{-i}$, in which case $c(s^A,s)=\infty$, or $s_{-i} = s^A_{-i}$, $s_i \neq A$, in which case

$$
\begin{aligned}
(27)\qquad c(s^A,s) &\underbrace{=}_{\text{by (23)}} g_i\left(\Upsilon_i(s_i, f_i(s^A))\right)\\
&\underbrace{=}_{\text{by (8)}} g_i\left(\max_{x_i\in S_i} U_i(x_i,(f_i(s^A))_{-i}) - U_i(s_i,(f_i(s^A))_{-i})\right)\\
&\underbrace{=}_{\text{by } f(s^A)=s^A} g_i\left(\max_{x_i\in S_i} U_i(x_i,(s^A)_{-i}) - U_i(s_i,(s^A)_{-i})\right)\\
&\underbrace{>}_{\substack{\text{by } s_i\neq A,\,(26),\\ g_i \text{ strictly increasing}}} 0.
\end{aligned}
$$

Therefore, the condition in Definition 6a is satisfied. □

Proof of Theorem 2.

Let P^ε be the Markov kernel of the process with $F = \{f_1,\ldots,f_n\}$. Define processes $P^{\varepsilon,f_1},\ldots P^{\varepsilon,f_n}$

as identical to P^{ε} except that $F = \{f_1\}, \ldots, F = \{f_n\}$ respectively. Note that

$$(28) \qquad P^{\varepsilon} = \sum_{i=1}^{n} \varphi(f_m) P^{\varepsilon, f_i}.$$

Let cost functions for $P^{\varepsilon, f_1}, \ldots P^{\varepsilon, f_n}$ be given by $c^{f_1}, \ldots, c^{f_n}$.

Assume s^A is GR-dominant. Lemma 4(i) implies that, for all $i \in N$, $f_m \in F$, $c_i^{f_m}$ is asymmetric towards A. Theorem N(i) then implies that, for $m = 1, \ldots, n$, c^{f_m} is asymmetric towards A. Given (28), that is P^{ε} is a convex combination of $P^{\varepsilon, f_1}, \ldots P^{\varepsilon, f_n}$, this implies that c is asymmetric towards A (Newton, 2020, Theorem 1). Theorem P(i) then implies that s^A is stochastically stable.

Assume s^A is strictly GR-dominant and, for all $i \in N$, g_i is strictly increasing and $f_i(s^A) = s^A$. Lemma 4(ii) implies that, for all $i \in N$, $f_m \in F$, $c_i^{f_m}$ is strictly asymmetric towards A. Theorem N(ii) then implies that, for $m = 1, \ldots, n$, c^{f_m} is strictly asymmetric towards A. Given (28), that is P^{ε} is a convex combination of $P^{\varepsilon, f_1}, \ldots P^{\varepsilon, f_n}$, this implies that c is strictly asymmetric towards A (Newton, 2019, Theorem 1). Theorem P(ii) then implies that s^A is uniquely stochastically stable.

□

Proof of Proposition 2.

Assume A is risk dominant. Step 1 of the Proof of Proposition 1 implies that s^A is GR-dominant. Therefore, the first part of Theorem 2 applies and s^A is stochastically stable.

□

References

Alós-Ferrer, C. and Netzer, N. (2010). The logit-response dynamics. *Games and Economic Behavior*, 68:413–427.

Battalio, R., Samuelson, L., and Van Huyck, J. (2001). Optimization incentives and coordination failure in laboratory stag hunt games. *Econometrica*, 69(3):749–764.

Block, H. and Marschak, J. (1959). Random orderings and stochastic theories of response. Working paper, Cowles Foundation for Research in Economics, Yale University.

Blume, L. E. (1993). The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387 – 424.

Blume, L. E. (1996). Population games. Working Papers 96-04-022, Santa Fe Institute.

Blume, L. E. (2003). How noise matters. *Games and Economic Behavior*, 44(2):251–271.

Cabrales, A., Nagel, R., and Armenter, R. (2007). Equilibrium selection through incomplete information in coordination games: an experimental study. *Experimental Economics*, 10(3):221–234.

Conlisk, J. (1993a). Adaptation in games: Two solutions to the crawford puzzle. *Journal of Economic Behavior & Organization*, 22(1):25–50.

Conlisk, J. (1993b). Adaptive tactics in games: Further solutions to the crawford puzzle. *Journal of Economic Behavior & Organization*, 22(1):51–68.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 215–242.

Crawford, V. (2007). Modeling behavior in novel strategic situations via level-k thinking. Conference slides - North American Economic Science Association Meetings.

Crawford, V. P. (1974). Learning the optimal strategy in a zero-sum game. *Econometrica: Journal of the Econometric Society*, pages 885–891.

Crawford, V. P. (2019). Experiments on cognition, communication, coordination, and cooperation in relationships.

Duffy, J. and Nagel, R. (1997). On the robustness of behaviour in experimental 'beauty contest' games. *The Economic Journal*, 107(445):1684–1700.

Foster, D. and Young, H. P. (1990). Stochastic evolutionary game dynamics. *Theoretical Population Biology*, 38:219–232.

Geanakoplos, J. and Gray, L. (1991). When seeing further is not seeing better. *Bulletin of the Santa Fe Institute*, 6(2):1–6.

Gill, D. and Prowse, V. (2016). Cognitive ability, character skills, and learning to play equilibrium: A level-k analysis. *Journal of Political Economy*, 124(6):1619–1676.

Goeree, J. K. and Holt, C. A. (2004). A model of noisy introspection. *Games and Economic Behavior*, 46(2):365–382.

Harsanyi, J. C. and Selten, R. (1988). A general theory of equilibrium selection in games. *MIT Press Books*, 1.

Heinemann, F., Nagel, R., and Ockenfels, P. (2004). The theory of global games on test: experimental analysis of coordination games with public and private information. *Econometrica*, 72(5):1583–1599.

Heller, Y. (2015). Three steps ahead. *Theoretical Economics*, 10(1):203–241.

Ho, T.-H. and Su, X. (2013). A dynamic level-k model in sequential games. *Management Science*, 59(2):452–469.

Hwang, S.-H., Lim, W., Neary, P., and Newton, J. (2018). Conventional contracts, intentional behavior and logit choice: Equality without symmetry. *Games and Economic Behavior*, 110:273 – 294.

Iijima, R. (2015). Iterated generalized half-dominance and global game selection. *Journal of Economic Theory*, 159:120–136.

Kandori, M., Mailath, G. J., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, 61(1):29–56.

Khan, A. and Peeters, R. (2014). Cognitive hierarchies in adaptive play. *International Journal of Game Theory*, 43(4):903–924.

Lim, W. and Neary, P. R. (2016). An experimental investigation of stochastic adjustment dynamics. *Games and Economic Behavior*, 100:208–219.

Mäs, M. and Nax, H. H. (2016). A behavioral study of "noise" in coordination games. *Journal of Economic Theory*, 162:195–208.

Matros, A. (2003). Clever agents in adaptive learning. *Journal of Economic Theory*, 111(1):110–124.

Mohlin, E. (2012). Evolution of theories of mind. *Games and Economic Behavior*, 75(1):299–318.

Morris, S., Rob, R., and Shin, H. S. (1995). p-dominance and belief potential. *Econometrica*, 63:145–157.

Myatt, D. P. and Wallace, C. (2003). Sophisticated play by idiosyncratic agents. *Journal of Evolutionary Economics*, 13(3):319–345.

Myerson, R. (1978). Refinements of the Nash equilibrium concept. *International Journal of Game Theory*, 7:73–80.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326.

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences*, 36(1):48–49.

Newton, J. (2017). Shared intentions: The evolution of collaboration. *Games and Economic Behavior*, 104:517 – 534.

Newton, J. (2018). Evolutionary game theory: A renaissance. *Games*, 9(2):31.

Newton, J. (2019). Combining strictly asymmetric choice rules. SSRN Working Paper Series 3433736.

Newton, J. (2020). Conventions under heterogeneous behavioural rules. *Review of Economic Studies*, (https://doi.org/10.1093/restud/rdaa063).

Peski, M. (2010). Generalized risk-dominance and asymmetric dynamics. *Journal of Economic Theory*, 145(1):216 – 248.

Proto, E., Rustichini, A., and Sofianos, A. (2019). Intelligence, personality, and gains from cooperation in repeated interactions. *Journal of Political Economy*, 127(3):1351–1390.

Robalino, N. and Robson, A. (2016). The evolution of strategic sophistication. *American Economic Review*, 106(4):1046–1072.

Rusch, H. (2019). The evolution of collaboration in symmetric 2×2-games with imperfect recognition of types. *Games and Economic Behavior*, 114:118–127.

Sáez-Martı, M. and Weibull, J. W. (1999). Clever agents in young's evolutionary bargaining model. *Journal of Economic Theory*, 86(2):268–279.

Sandholm, W. H. (2010). *Population games and evolutionary dynamics*. Economic learning and social evolution. Cambridge, Mass. MIT Press.

Selten, R. (1991). Anticipatory learning in two-person games. In *Game equilibrium models I*, pages 98–154. Springer.

Stahl, D. O. (1993). Evolution of smartn players. *Games and Economic Behavior*, 5(4):604–617.

Stahl, D. O. (1999). Evidence based rules and learning in symmetric normal-form games. *International Journal of Game Theory*, 28(1):111–130.

Stahl, D. O. (2000). Rule learning in symmetric normal-form games: theory and evidence. *Games and Economic Behavior*, 32(1):105–138.

Stahl, D. O. and Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of Economic Behavior and Organization*, 25(3):309–327.

Stennek, J. (2000). The survival value of assuming others to be rational. *International Journal of Game Theory*, 29(2):147–163.

Tang, F.-F. (2001). Anticipatory learning in two-person games: some experimental results. *Journal of economic behavior & organization*, 44(2):221–232.

Van Huyck, J. B., Battalio, R. C., and Beil, R. O. (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, pages 234–248.

Young, H. P. (1993a). The evolution of conventions. *Econometrica*, 61(1):57–84.

Young, H. P. (1993b). An evolutionary model of bargaining. *Journal of Economic Theory*, 59(1):145 – 168.