Appendix B. Proofs of propositions in chapter 5

Proof of Proposition 5-1. Proving the second part of proposition 5-1 –i.e. that the asymptotic behaviour of the N-CBR model is independent of the decision-making algorithm employed by each player *i* when she has not yet explored every action available to her in a *similar* situation– is straightforward, since this is a transient situation. Given the definition of the set of different states of the world possibly perceived by any player, the trembling hands noise guarantees that sooner or later every possible state of the world perceived by any player will happen infinitely often. The trembling hands noise also guarantees that every player will choose every possible action available to her in any given situation. Thus, sooner or later, every player will have selected every action available to her in every possible state of the world she can perceive (i.e. every action available to player *i* will be represented in her set of cases C_i, for every state of the world possibly perceived by *i*). Therefore, sooner or later, no player will be using the decision-making algorithm that the second part of proposition 5-2 refers to, so the asymptotic behaviour of the model is independent of such algorithms.

The following proves part 1 of proposition 5-1, i.e. that if every player has a common perception of the state of the world, then the asymptotic behaviour of the N-CBR process is independent of the specific structure of the perceived state of the world. The previous paragraph demonstrates that sooner or later the state of the system in the N-CBR model is fully characterised by every player's set of most recent cases that occurred in every possible perceived state of the world for each one of the actions available to her. Thus, this second proof (which refers to the asymptotic behaviour of the system) assumes that every player has already selected every action available to her at least once in every possible state of the world she can perceive. Consider the following two points:

- The assumption that players have a common perception of the state of the world implies that all players perceive that any particular state of the world has occurred in exactly the same time-steps. In other words, all players would unanimously agree or disagree with any statement of the form "The situations lived in time-steps $\{x, y, \ldots, z\}$ looked all similar to me (i.e. they correspond to the same perceived state of the world)".

- Note also that the decision made by each player i in any particular situation is only affected by decisions (made by all players) that took place in a previous *similar* situation (i.e. having perceived the same state of the world).

Thus, one can view the dynamics of the whole model (where players can perceive various different states of the world) as a collection of parallel dynamic processes, each of them corresponding to one specific state of the world (perceived by all players at once). The dynamics observed for each individual perceived state of the world are governed by the same decision-making processes and are independent of each other. Each of these individual threads, if observed on its own, induces the same dynamics that one would observe in a model where players cannot distinguish between different states of the world. The following table illustrates this interpretation with an example.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
SW_t	sw_3	sw_1	sw_4	sw_3	sw_2	sw_3	sw_4	sw_4	sw_1	sw_2	sw_1	sw_3	sw_2	sw_4	sw_1
THREAD $SW = sw_1$		1							2		3				4
THREAD $SW = sw_2$					1					2			3		
THREAD $SW = sw_3$	1			2		3						4			
THREAD $SW = sw_4$			1				2	3						4	

where SW_t is the random variable that denotes the state of the world perceived by every player at time-step t, sw_i are particular values of that variable, and the numbers on coloured backgrounds inside the table indicate the number of times that the corresponding state of the world has been visited.

Let $T_n^{(sw)}$ be the state of the thread $\{SW = sw\}$ (where the perceived state of the world is sw), defined by the payoffs each player obtained the last time that she selected each of the actions available to her having observed state of the world sw, when state of the world sw has been observed n times. It is clear then that the sequence of random variables $\{T_n^{(sw)}\}_{n\geq 1}$ (for any fixed sw) corresponds to a model where players cannot distinguish between different states of the world. Following the reasoning presented in the first paragraph of section 5.7, it is also straightforward to show that $\{T_n^{(sw)}\}_{n\geq 1}$ can be formulated as a uni-reducible Markov chain, which has a unique limiting distribution (Janssen and Manca, 2006, Corollary 5.2, pg. 117). Finally, it should also be apparent that all threads have the same limiting distribution:

$$\lim_{n\to\infty} \mathbf{Pr}(T_n^{(i)} = \alpha) = \lim_{n\to\infty} \mathbf{Pr}(T_n^{(j)} = \alpha) \quad \forall i, j$$

For clarity of notation, let $\{T_n\}_{n\geq 1}$ denote the sequence of states corresponding to a model where players cannot distinguish different states of the world. Thus,

$$\lim_{n\to\infty} \mathbf{Pr}(T_n^{(i)} = \alpha) = \lim_{n\to\infty} \mathbf{Pr}(T_n^{(j)} = \alpha) = \lim_{n\to\infty} \mathbf{Pr}(T_n = \alpha) \qquad \forall i, j$$

The fact that remains to be proven is that the overall dynamics of the model (i.e. the ensemble of threads) also show the same limiting distribution as the individual threads. To show that, let X_t denote the state of the thread corresponding to the state of the world observed at time t. Formally:

$$X_t = \left\{T_{N_i(t)}^{(i)} : SW_t = i\right\}$$

where $N_i(t)$ denotes the number of times that the event $\{SW_t = i\}$ has occurred up until time-step t. Formally: $N_i(t) = \#\left\{ k \in \{1,...,t\} : SW_k = i\right\}$

With this notation, the proof of the second part of proposition 5-2 will be concluded once it is demonstrated that:

$$\lim_{t\to\infty} \mathbf{Pr}(X_t = \alpha) = \lim_{t\to\infty} \mathbf{Pr}(T_t = \alpha)$$

The following, which is conditioned to a set of (arbitrary) initial conditions, concludes the proof.

$$\lim_{t\to\infty} \mathbf{Pr}(X_t = \alpha) = \lim_{t\to\infty} \sum_i \mathbf{Pr}(\{SW_t = i\} \ \& \ \{T_{N_i(t)}^{(i)} = \alpha\}) =$$

$$= \lim_{t\to\infty} \sum_i \mathbf{Pr}(T_{N_i(t)}^{(i)} = \alpha \mid SW_t = i) \cdot \mathbf{Pr}(SW_t = i)$$

It has been argued previously that states of the world are visited infinitely often, thus:

$$\lim_{t\to\infty} \mathbf{Pr}(T_{N_i(t)}^{(i)} = \alpha \mid SW_t = i) = \lim_{n\to\infty} \mathbf{Pr}(T_n^{(i)} = \alpha) = \lim_{n\to\infty} \mathbf{Pr}(T_n = \alpha)$$

(Regardless of the set of (arbitrary) initial conditions)

and it is also clear that $\sum_i \mathbf{Pr}(SW_t = i) = 1 \qquad \forall t$

Using the two results above the first part of proposition 5-1 is finally proved:

$$\lim_{t\to\infty} \mathbf{Pr}(X_t = \alpha) = \lim_{t\to\infty} \sum_i \mathbf{Pr}(T^{(i)}_{N_i(t)} = \alpha \mid SW_t = i) \cdot \mathbf{Pr}(SW_t = i) = \lim_{t\to\infty} \mathbf{Pr}(T_t = \alpha)$$

Proof of Proposition 5-2. As argued in the proof of proposition 5-1, sooner or later, every player will have selected every action available to her in every possible state of the world she can perceive (i.e. every action available to player *i* will be represented in her set of cases C_i, for every state of the world possibly perceived by *i*). Thus, sooner or later, the state of the system in the N-CBR model is fully characterised by every player's set of most recent payoffs she obtained for each one of the actions available to her in every possible state of the world she can perceive. The model thus defined is a finite-state irreducible aperiodic discrete-time Markov chain, which is denoted P^ε. Let P^0 be the Markov process P^ε when $\varepsilon = 0$ and all players have explored all their available actions for every possible state of the world they can perceive. Note that P^0 is generally reducible.

The proof rests on two arguments. The first argument, which is an immediate application of theorem 4 in Young (1993), is that every stochastically stable state is a recurrent state of P^0 (i.e. the model without noise). The second argument is that the *outcome* (i.e. the set of decisions made by players) that is induced by any recurrent *state* of P^0 is necessarily individually rational. The following proves an alternative (but equivalent) formulation of the second argument: if state *x* in P^0 induces an outcome that is not individually rational, then *x* is a transient state of P^0. We will prove this second argument by showing that if state *x* induces an outcome that is not individually rational, then *x* will never be revisited.

Let *A* be one of the players who has received a payoff below her maximin $Maximin_A$ in the outcome induced by state *x*, and let sw_A be the state of the world perceived by *A* in state *x*. Let *a* be the action that *A* chose in state *x*, and $p_x(A, a)$ be the payoff that *A* had obtained the previous time she had perceived sw_A and selected action a; this payoff $p_x(A, a)$ is part of the definition of x. Note that a necessary condition for x to be revisited is that player A perceives sw_A again, and also that the payoff that A has obtained the previous time she has perceived sw_A and selected action a is $p_x(A, a)$. This can never be the case for the following argument:

1. The fact that player A selected action a in state x implies that $p_x(A, a) \geq Maximin_A$. In more informal terms, the payoff player A believed she would obtain by selecting action a (having observed state of the world sw_A) was the maximum over all her possible actions, and therefore it was necessarily no less than $Maximin_A$.
2. Player A obtained a payoff strictly below her $Maximin_A$ when, after having perceived state of the world sw_A, she selected action a. Thus, from then onwards she will remember that the last time she selected action a having observed state of the world sw_A, she obtained a payoff strictly below $Maximin_A$.
3. There is at least one action that gives player A a payoff no less than $Maximin_A$ regardless of the actions of her counterparts. When perceiving state of the world sw_A again, player A will always select this (maximin) action over action a. Thus, player A will never update her belief that selecting action a when she perceives state of the world sw_A will give her a payoff below $Maximin_A$.

State x required player A to believe that selecting action a would give her a payoff no less than $Maximin_A$. Thus, state x cannot be revisited, and this fact concludes the proof.