Chapter 3
Adaptive Dynamics in Traffic Games

The equilibrium concepts discussed in the previous chapters were eminently static, with traffic modeled as a steady state that results from the adaptive behavior of drivers. These models assume implicitly the existence of a hidden mechanism that leads to equilibrium, though they are stated directly in the form of equilibrium equations which are not tied to a specific adaptive dynamics. Furthermore, they consider a non-atomic framework that ignores individual drivers and uses continuous variables to represent aggregate flows. In this chapter we discuss a dynamical model for the adaptive behavior of finitely many drivers in a simple network and we study its convergence towards a steady state.

Empirical evidence of some form of adaptive behavior has been reported in [8, 49, 68], based on experiments and simulations of different discrete time adaptive dynamics, though it has been observed that the steady states attained may differ from the standard equilibria and may even depend on the initial conditions. Also, empirical support for the use of discrete choice models in the context of games was given in [57], while the convergence of a class of finite-lag adjustment procedures was established in [23, 24, 29]. On a different direction, several continuous time dynamics describing plausible adaptive mechanisms that converge to Wardrop equilibrium were studied in [41, 67, 73], though these models are of an aggregate nature and are not directly linked to the behavior of individual drivers.

Learning and adaptive behavior in repeated games have been intensively explored in the last decades (see [42, 79]). Instead of studying aggregate dynamics in which the informational and strategic aspects are unspecified, the question is considered from the perspective of individual players that make decisions day after day based on information derived from past observations. The accumulated experience is summarized in a state variable that determines the strategic behavior of players through a certain stationary rule. At this level, the most prominent adaptive procedure is fictitious play, studied by Brown [22] in the early 50s (see also [63]), which assumes that at each stage players choose a best reply to the
observed empirical distribution of past moves of their opponents. A variant called smooth fictitious play for games with perturbed payoffs and reminiscent of Logit random choice is discussed in [42, 48].

The assumption that players are able to record the past moves of their opponents is very stringent for games involving many players with limited information. A milder assumption is that players observe only the outcome vector, namely the payoff obtained at every stage and the payoff that would have resulted if a different move had been played. Procedures such as no-regret [44, 45], exponential weight [40], and calibration [38], deal with such limited information contexts assuming that players adjust their behavior based on statistics of past performance. Eventually, adaptation leads to configurations where no player regrets the choices he makes. These procedures, initially conceived for the case when the complete outcome vector is available, were adapted to the case where only the realized payoff can be observed [6, 38, 42, 45]. The idea is to use the observed payoffs to build an unbiased estimator of the outcome vector to which the initial procedure is applied.

Although these procedures are flexible and robust, the underlying rationality may still be too demanding for games with a large number of strategies and poorly informed players. This is the case for traffic where a multitude of small players make routing decisions with little information about the strategies of other drivers nor the actual congestion in the network. A simpler adaptive rule which relies only on the past sequence of realized moves and payoffs is reinforcement dynamics [7, 10, 19, 33, 53, 61]: players select moves proportionally to a propensity vector built by a cumulative procedure in which the current payoff is added to the component played, while the remaining components are kept unchanged. This mechanism is related to the replicator dynamics [61] and its convergence has been established for zero-sum 2-player games as well as for some games with unique equilibria [10, 53]. We should also mention here the mechanism proposed in [19], which uses an averaging rule with payoff dependent weights.

A similar idea is considered in this chapter. The model assumes that each player has a prior perception or estimate of the payoff performance for each possible move and makes a decision based on this rough information by using a random choice rule such as Logit. The payoff of the chosen alternative is then observed and is used to update the perception for that particular move. This procedure is repeated day after day, generating a discrete time stochastic process which we call the learning process. The basic ingredients are therefore: a state parameter; a decision rule from states to actions; and an updating rule on the state space. This structure is common to many procedures in which the incremental information leads to a change in a state parameter that determines the current behavior through a given stationary map. However, the specific dynamics considered here are structurally different from those studied previously, while preserving the qualitative features of probabilistic choice and sluggish adaptation [79, Section 2.1]. Although players observe only their own payoffs, these values are affected by everybody else’s choices revealing information on the game as a whole. The basic