Chapter 66

LEARNING AND EQUILIBRIUM IN GAMES

COLIN F. CAMERER

Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125, USA
e-mail: camerer@hss.caltech.edu

TECK H. HO

The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6366, USA
e-mail: hoteck@wharton.upenn.edu

JUIN-KUAN CHONG

National University of Singapore, Lower Kent Ridge Road, Singapore 192610

1. Introduction

In the last ten years theory (e.g., Fudenberg and Levine, 1998) and empirical data fitting have provided many ideas about how equilibria arise in games or markets. This short chapter describes a very general approach to learning in games: "experience-weighted attraction" (EWA) learning. This approach strives to explain, for every choice in an experiment, how that choice arose from players' previous behavior and experience, using a general model which can be applied to most games with minimal customization and which predicts well out of sample. Sophisticated EWA includes important equilibrium concepts and many other learning models (simple reinforcement, Cournot, fictitious play, weighted fictitious play) as special cases (see Camerer and Ho, 1999; Ho, Camerer and Chong, 2001, and cited references for details). The model therefore allows "one-stop shopping" for learning about the latest statistical comparisons of many different learning and equilibrium models (see Camerer, 2002, Chapter 6 for more details). The model can also be adapted to field applications in which strategies and payoffs are often poorly-specified (e.g., it has been used successfully to predict actual consumer choices of products like ice cream, see Ho and Chong, 1999).

2. Adaptive EWA and Other Learning Models

Notation: Denote player i's jth strategy by s_i^j and the other player(s)' strategy by s_{-i}^k. The strategy actually chosen in period t is $s_i(t)$. Player i's payoff for choosing s_i^j in period t is $\pi(s_i^j, s_{-i}^k(t))$. Like most learning theories, EWA assumes each strategy has a numerical measure, called an attraction $A_i^j(t)$. The model also has an experience weight, $N(t)$. The variables $N(t)$ and $A(a, t)$ begin with prior values (estimated from

Handbook of Experimental Economics Results, Volume 1
Copyright © 2008 Elsevier B.V. All rights reserved
DOI: 10.1016/S1574-0722(07)00066-2

the data or specified from a model of first-period play) and are updated each period. The rule for updating attraction sets $A_i^j(a,t)$ to be the sum of a depreciated, experience-weighted previous attraction $A_i^j(a,t-1)$ plus the (weighted) payoff from period t, normalized by the updated experience weight (argument a stands for adaptive learning):

$$A_i^j(a,t) = \frac{\phi \cdot N(t-1) \cdot A_i^j(a,t-1) + [\delta + (1-\delta) I(s_i^j, s_i(t))] \cdot \pi_i(s_i^j, s_{-i}(t))}{N(t)} \tag{2.1}$$

where indicator variable $I(x,y)$ is 1 if $x = y$ and 0 otherwise. The experience weight is updated by $N(t) = \phi(1-\kappa)N(t-1) + 1$. In Bayesian models (belief special cases with Dirichlet priors), $N(0)$ is the strength of prior beliefs, in terms of units of observation. In general, $N(t)$ approaches the steady-state value $\frac{1}{1-\phi\cdot(1-\kappa)}$ value, it steadily rises, capturing an increase in the weight placed on previous attractions and a (relative) decrease in the impact of recent observations, so that learning slows down. [In practice, assuming $N(0) = 1$ or imposing the restriction $N(t)\frac{1}{1-\phi\cdot(1-\kappa)}$ save degrees of freedom and impair fit very little.]

Attractions are mapped into choice probabilities using an exponential logit rule (other functional forms fit about equally well; Camerer and Ho, 1998):

$$P_i^j(a,t+1) = \frac{e^{\lambda \cdot A_i^j(a,t)}}{\sum_{k=1}^{m_i} e^{\lambda \cdot A_i^k(a,t)}}. \tag{2.2}$$

The key parameters are δ, ϕ, and κ (which are generally assumed to be in the [0, 1] interval). When $\kappa = 0$, the attractions are weighted averages of lagged attractions and payoff reinforcements (with weights $\phi\cdot N(t-1)/(\phi\cdot N(t-1)+1)$ and $1/(\phi\cdot N(t-1)+1)$). When $\kappa = 1$, $N(t) = 1$, the attractions are accumulations of previous reinforcements rather than averages (i.e., $A_i^j(a,t) = \phi\cdot A_i^j(a,t-1)+[\delta+(1-\delta)\cdot I(s_i^j, s_i(t))]\cdot\pi_i(s_i^j, s_{-i}(t))$). In the logit model, the *differences* in strategy attractions entirely determine their choice probabilities. When κ is high, the attractions can grow furthest apart over time, making choice probabilities closer to zero and one. We therefore interpret k as an index of "commitment" or cumulation. It seems related to a distinction in machine learning between exploration (trying different strategies to see which is best) and exploitation (locking in to the strategy which has worked best). High values of κ correspond to quicker exploitation. The parameter ϕ represents the rate at which old experience is discounted relative to new, a measure of sensitivity to change. The most important parameter, δ, is the weight on foregone payoffs relative to realized payoffs, "consideration" or imagination.

Triples of parameter values δ, ϕ, κ represent specific learning rules, which can be shown in a three-dimensional cube (see Figure 1). Simple algebra shows that certain corners and vertices of the cube correspond to extreme special cases which are historically significant. The vertex $\delta = 1$, $\kappa = 0$ corresponds, surprisingly, to weighted fictitious play models in which players form beliefs based on past observation of others, and

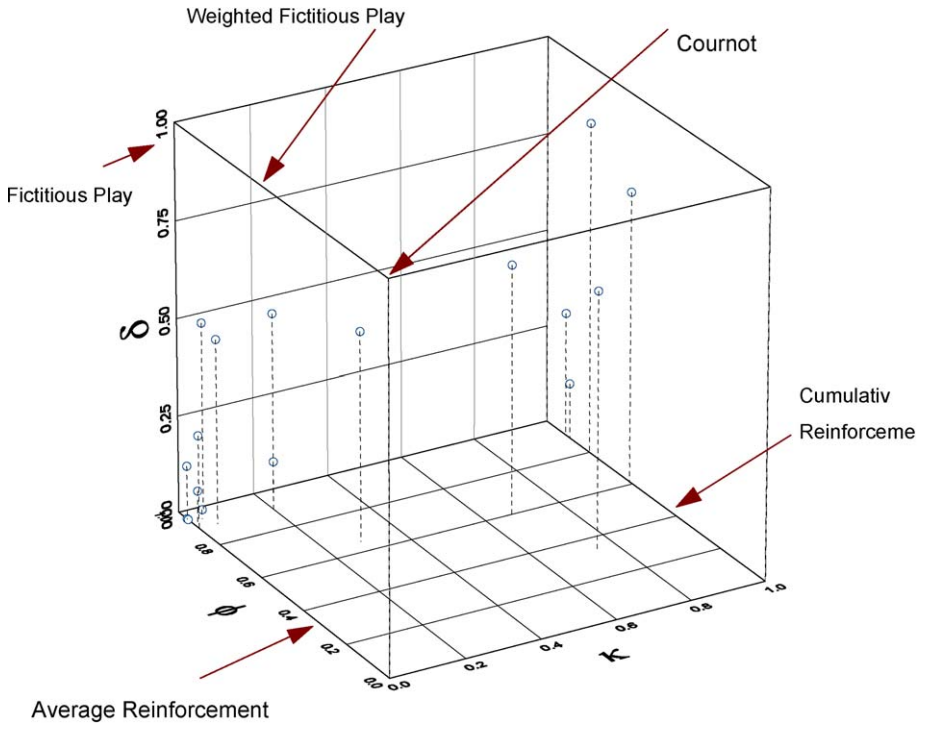


Figure 1.

choose best responses given their beliefs. The corners $\phi = 0$ and $\phi = 1$ correspond to Cournot best-response dynamics and fictitious play, respectively. Reinforcement models in which only chosen strategies are reinforced according to their payoffs correspond to vertices in which $\delta = 0$, and $\kappa = 1$ (cumulative reinforcement) or $\kappa = 0$ (averaged reinforcement). Interior configurations of parameter values incorporate both the intuition behind reinforcement learning, that realized payoffs weigh most heavily ($\delta < 1$), and the intuition implicit in belief learning, that foregone payoffs matter too ($\delta > 0$). The vertex $\delta = 1$, $\kappa = 0$ represents "cumulative" best-response learning. Though that vertex has never been studied, its parameter restrictions fit best in three coordination games (Ho, Camerer and Chong, 2001), which shows the advantage of hybridizing the responsiveness of belief learning (high δ) with the cumulation of some reinforcement models (high κ).

The cube shows that contrary to popular belief for many decades, reinforcement and belief learning are not fundamentally different approaches to learning. Instead, they are simply two extreme configurations on opposite edges of a three-dimensional cube. That is, belief learning (of the fictitious play variety) is a kind of generalized reinforcement in which unchosen strategies are reinforced as strongly as chosen ones. Parameter

estimates in a wide variety of experimental data sets also show the empirical advantages of hybridizing the three features of learning. Figure 1 also shows estimates of the three parameters in 20 different studies (Ho, Camerer and Chong, 2001). The estimation maximizes the likelihood function using each data point (there is no averaging across individuals or blocks of trials). About 70% of the data are used to estimate best-fitting parameter values, and those values are used to forecast data in the remaining 30% of the data to be sure the model is not fitting well by overfitting. While this procedure is different than fixing a set of parameter values and simulating a priori an entire sample path, there is no evidence that in explaining data from games the maximum-likelihood and simulated-path methods yield different results on relative fit of models (see Camerer, 2002, Chapter 6).

Each point in Figure 1 is a triple of estimates. Most points are sprinkled throughout the cube, rather than at the extreme vertices mentioned in the previous paragraph, although some (generally from games with mixed-strategy equilibria) are near the averaged reinforcement corner $\delta = 0, \kappa = \phi = 1$.

Parameter estimates are generally significantly inside the interior of the cube, rather than near the vertices. That means the general EWA specification fits better than cumulative reinforcement (with $\kappa = 1$) in 27 of 31 cases, and better than belief learning in 25 of 27 data sets (penalizing for free parameters or predicting out-of-sample). Given the plausibility of the hybrid EWA model, the psychological interpretability of its parameters, the efficiency of searching for optimal parameters in the entire cube rather than along a single vertex or corner, and its demonstrated superiority in more than 90% of comparisons in 25–30 data sets, it is hard to think of a good reason to continue to focus only on extreme special cases rather than EWA.

One concern about a model like EWA is that it has "too many" parameters, and the parameters vary across games (so it might be difficult to guess what values would be in a new game). In Ho, Camerer and Chong (2001) both problems are solved by substituting functions of experience for free parameters. (For example, ϕ is a "change-detection" function which dips down below zero, discarding old information, if an opponent's strategies change dramatically.) This "functional EWA" (or fEWA) model has only one free parameter (λ) and is hence more parsimonious than most reinforcement and belief learning models. Furthermore, the functional values which fEWA generates tend to be close to estimated values across games, reducing cross-game unpredicted variation. Finally, Ho et al. propose a measure of the theory's "economic value" – if players followed theory recommendations, by best-responding to theory forecasts of others' behavior rather than making the choices they did, how much more money would they have earned? Across seven data sets, EWA and fEWA add the most economic value by this measure, compared to general belief, reinforcement, and QRE models. To guard against the possibility that the original model was overfit, Ho et al also collected three more data sets after their first draft was written and found that performance on those new data was comparable to the earlier ones.

3. Sophisticated EWA and Equilibrium Models

The EWA model presented is a simplification (as are all the other adaptive models) because it does not permit players to anticipate learning by others (cf. Selten, 1986). Omitting anticipation logically implies that players do not use information about the payoffs of other players, and that whether players are matched together repeatedly or randomly re-matched should not matter. Both of the latter implications are unintuitive and have proved false in experiments, and there is direct evidence for anticipatory learning also.

In Camerer, Ho, and Chong (2002a, 2002b) we propose a simple way to include "sophisticated anticipation" by some players that others are learning, using two parameters. We assume a fraction of players are sophisticated. Sophisticated players think that a fraction $(1 - \alpha')$ of players are adaptive and the remaining fraction α' of players are sophisticated like themselves. They use the adaptive EWA model to forecast what the adaptive players will do, and choose strategies with high expected payoffs given their forecast.

All the adaptive models discussed above (EWA, reinforcement, weighted fictitious play) are special cases of sophisticated EWA with $\alpha = 0$. The assumption that sophisticated players think some others are sophisticated, creates a small whirlpool of recursive thinking which implies that quantal response equilibrium (QRE; McKelvey and Palfrey, 1998) and hyperresponsive QRE (Nash) equilibrium, are special cases of sophisticated EWA. Our specification also shows that equilibrium concepts combine two features which are empirically and psychologically separable: "social calibration" (accurate guesses about the fraction of players who are sophisticated, $(\alpha > \alpha')$ and full sophistication $(\alpha = 1)$). Psychologists have identified systematic departures from social calibration called "false" uniqueness or overconfidence $(\alpha > \alpha')$ and "false" consensus or curse of knowledge $(\alpha > \alpha')$.

Formally, adaptive EWA learners follow the updating equations above. Sophisticated players have attractions and choice probabilities specified as follows (where arguments a and s denote adaptive and sophisticated, respectively):

$$A_i^j(s,t) = \sum_k \left[(1-\alpha')\cdot P_{-i}^k(a,t+1) + \alpha' P_{-i}^k(s,t+1)\right]\cdot \pi_i\left(s_i^j, s_{-i}^k\right), \tag{3.1}$$

$$P_i^j(s,t+1) = \frac{e^{\lambda\cdot A_i^j(s,t)}}{\sum_{k=1}^{m_i} e^{\lambda\cdot A_i^k(s,t)}}. \tag{3.2}$$

The sophisticated model has been applied to experimental data from 10-period p-beauty contest games (Ho, Camerer, and Weigelt, 1998). In these games, seven subjects choose numbers in [0, 100] simultaneously. The subject whose number is closest to p times the average (where $p = .7$ or .9) wins a fixed prize. Subjects playing for the first time are called "inexperienced"; those playing another 10-period game (with a different p) are called "experienced."

Table 1
Parameter estimates for p-beauty contest game

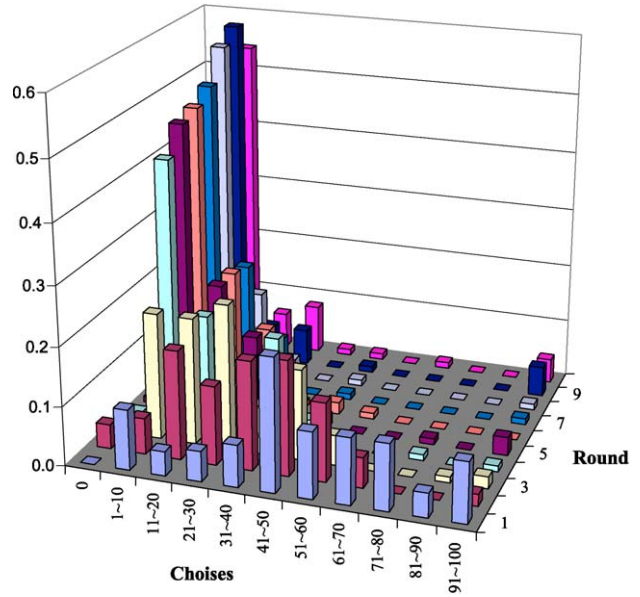
	Inexperienced subjects		Experienced subjects	
	Sophisticated EWA	Adaptive EWA	Sophisticated EWA	Adaptive EWA
ϕ	**0.436**	0.000	**0.287**	0.220
δ	**0.781**	0.900	**0.672**	0.991
κ	**1.000**	1.000	**0.927**	1.000
$N(O)$	**0.253**	0.000	**0.000**	0.887
α	**0.236**	*0.000*	**0.752**	*0.000*
α'	**0.000**	*0.000*	**0.412**	*0.000*
LL (in sample)	**−2095.32**	−2155.09	**−1908.48**	−2128.88
LL (out of sample)	**−968.24**	−992.47	**−710.28**	−925.09

Table 1 reports results and parameter estimates. For inexperienced subjects, adding sophistication to adaptive EWA improves log likelihood (LL) substantially both in- and out-of-sample. The estimated fraction of sophisticated players is $\hat{\alpha} = .236$ and their estimated perception $\hat{\alpha}' = 0$. The consideration parameter δ is estimated to be .781.

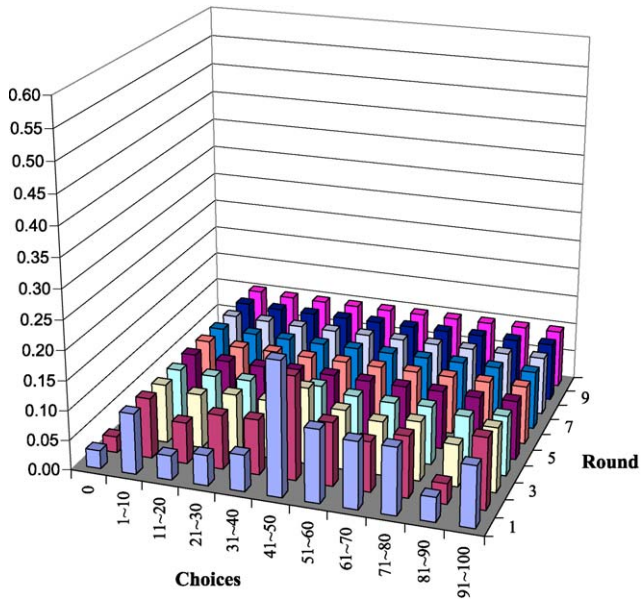
Experienced subjects show a larger improved fit from adding sophistication, and a larger estimated proportion, $\hat{\alpha} = .752$. (Their perceptions are again too low, $\hat{\alpha}' = .413$, showing a degree of overconfidence.) The increase in sophistication due to experience reflects a kind of "cross-period" learning which is similar to rule learning (Stahl, 2003).

Figure 2a shows actual choice frequencies for experienced subjects across the ten periods. Figures 2b–2d show predicted frequencies for choice reinforcement, weighted fictitious play, and sophisticated EWA. Figure 2b shows that reinforcement learns far too slowly because only one player wins each period and the losers get no reinforcement. [The reinforcement model in Roth and Erev (1995) has a simpler problem in games with proposer competition, which they circumvent by reinforcing ex post winning strategies in a way much like EWA updating.] Figure 2c shows that belief models with low values of ϕ, update beliefs very quickly but do not capture anticipatory learning, in which subjects anticipate that others will best-respond and leapfrog ahead. As a result, the frequency of low choices (1–10) predicted by belief learning only grows from 20% in period 5 to 35% in period 10, while the actual frequencies grow from 40% to 55%. Adding sophistication (Figure 2d) captures those actual frequencies quite closely.

An important implication of sophistication we are exploring in current research is "strategic teaching" (Camerer, Ho, and Chong, 2002a, 2002b): Sophisticated players who are matched with the same players repeatedly may have an incentive to "teach" adaptive players, choosing strategies with poor short-run payoffs which will change what adaptive players do, in a way that benefits the sophisticated player (e.g., Fudenberg and Levine, 1989; Watson, 1993). Strategic teaching provides a learning-based foundation to theories of reputation formation and appears to fit better than type-based equilibrium approaches (even allowing for quantal response) in experimental data on

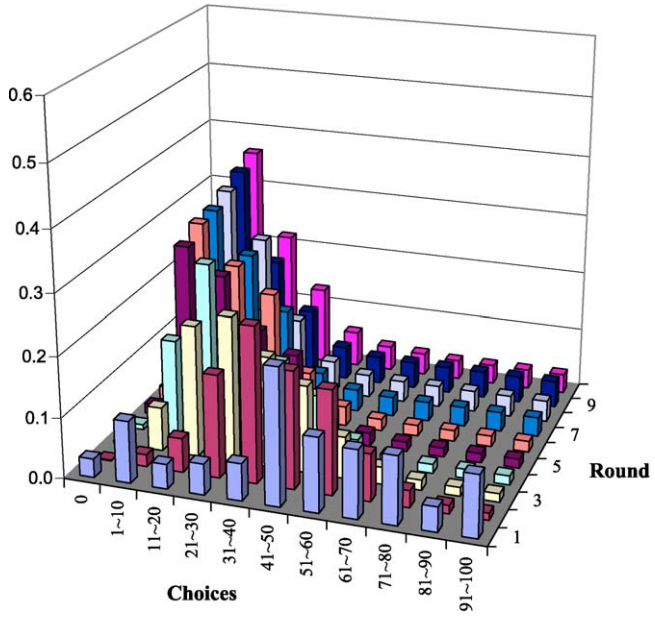


(a)

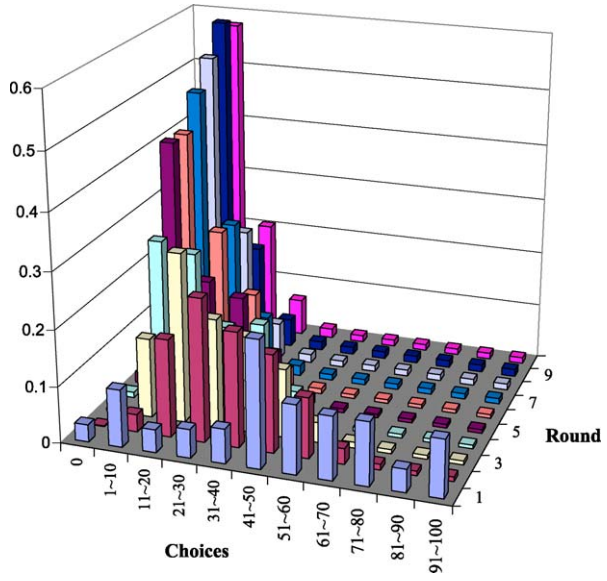


(b)

Figure 2. (a) Actual choice frequencies for experienced subjects. (b) Reinforcement model frequencies for experienced subjects. (c) Belief learning model frequencies for experienced subjects. (d) Sophisticated EWA model frequencies for experienced subjects.



(c)



(d)

Figure 2. (*continued*)

repeated entry deterrence (chain-store; Jung, Kagel, and Levin, 1994) and trust games (Camerer and Weigelt, 1988).

References

Camerer, Colin.F. (2002). "Behavioral Game Theory: Experiments on Strategic Interaction". Princeton University Press, Princeton.

Camerer, Colin F., Ho, Teck-Hua (1998). "EWA learning in normal-form games: Probability rules, heterogeneity and time-variation". Journal of Mathematical Psychology 42, 305–326.

Camerer, Colin F., Ho, Teck-Hua (1999). "Experience-weighted attraction learning in normal-form games". Econometrica 67, 827–874.

Camerer, Colin F., Weigelt, Keith (1988). "An experimental test of a sequential equilibrium reputation model". Econometrica 56, 1–36.

Camerer, Colin F., Ho, Teck-Hua, Chong, Juin-Kuan (2002a). "Sophisticated experience-weighted attraction learning and strategic teaching in repeated games". Journal of Economic Theory 104, 137–188.

Camerer, Colin F., Ho, Teck-Hua, Chong, Juin-Kuan (2002b). "Strategic teaching and equilibrium models of repeated trust and entry games". Caltech working paper, http://www.hss.caltech.edu/camerer/camerer.html.

Fudenberg, Drew, Levine, David (1998). "The Theory of Learning in Games". MIT Press, Cambridge.

Fudenberg, Drew, Levine, David (1989). "Reputation and equilibrium selection in games with a patient player". Econometrica 57, 759–778.

Ho, Teck-Hua, Chong, Juin-Kuan (1999). "A parsimonious model of SKU choice: Familiarity-based reinforcement and response sensitivity". Wharton Department of Marketing.

Ho, Teck-Hua, Camerer, Colin, Chong, Juin Kuan. (2001). "Economic value of EWA Lite: A functional theory of learning in games". Unpublished, http://www.hss.caltech.edu/camerer/camerer.html.

Ho, Teck-Hua, Camerer, Colin, Weigelt, Keith (1998). "Iterated dominance and iterated best-response in p-beauty contests". American Economic Review 88 (4), 947–969.

Jung, Yun Joo, Kagel, John H., Levin, Dan (1994). "On the existence of predatory pricing: An experimental study of reputation and entry deterrence in the chain-store game". RAND Journal of Economics 25, 72–93.

McKelvey, Richard D., Palfrey, Thomas R. (1998). "Quantal response equilibria for extensive form games". Experimental Economics 1, 9–41.

Roth, Alvin E., Erev, I. (1995). "Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term". Games and Economic Behavior, 164–212.

Selten, Reinhard (1986). "Anticipatory learning in 2-person games". University of Bonn Discussion Paper Series B.

Stahl, Dale (2003). "Sophisticated learning and learning sophistication". Working paper, University of Texas at Austin.

Watson, Joel (1993). "A 'reputation' refinement without equilibrium". Econometrica 61, 199–205.