NSF Fellowship:*

Graduate Research Plan Statement

Modibo K. Camara

October 24, 2016

This proposal addresses the problem of effective inference in the presence of like-minded others. While the literature on learning in games has made substantial progress in introducing various learning algorithms, these tend to suffer from a few widespread issues that limit their applicability. Foremost among these is the failure of incentive compatibility: if a learning algorithm is collectively adopted, it will not be individually optimal. I describe two approaches to overcome this failure. The first imposes an equilibrium condition on the adoption of learning algorithms themselves. The second reduces competition in the full space of learning algorithms to competition in computing power.

1 The Interactive Learning Problem

Consider the inference problem faced by high-frequency trading (HFT) algorithms. The trader relies on algorithms that can function well under the high frequency regime in which today's financial exchanges typically operate – well beyond human reaction speeds and perceptive capacity. The algorithm must achieve the standard single-agent learning objective of effective inference about the world around it, conditional on its own behavior. Because other trading algorithms that are both acting and learning in its presence, the inference may be highly non-stationary. The algorithm potentially knows nothing about the number of agents operating in its midst, let alone their incentives. The algorithm will find it difficult to disentangle fundamental from social dynamics, and must treat them jointly.

This is one case of a more general problem, learning in the context of game theory. Agents learn about other agents through their behavior, and in some cases, may act to influence said learning (i.e. strategic teaching). Strategies are thought of as an optimal repeated game strategy under a subjective prior belief distribution (Kalai and Lehrer 1993). The learning

*I would like to thank Ehud Kalai for his valuable comments and feedback. All errors are my own.

algorithm is fully embedded in the prior belief; that is, the choice of one's default prior corresponds to the choice of how to learn[1].

As Shoham et al. (2007) observe, this problem has been of interest at least since the advent of game theory itself (see Brown (1951) on fictitious learning). In economics, the literature has been framed largely as a justification for particular equilibrium concepts, so that learning is proposed as a process that leads toward said equilibrium. This literature is surveyed in Fudenberg and Levine (1998) and Young (2004), extending beyond Nash equilibrium to support correlated equilibrium (Foster and Vohra 1997) and self-confirming equilibrium (Fudenberg and Levine 1993) as the outcome of learning. Many of the algorithms developed capture relatively simple dynamics, sufficient for their purpose but requiring the user to adopt stationarity assumptions that do not apply to the environment in which they are situated. In computer science, beyond providing a computational method for finding equilibria, the literature has been used as a basis for multi-agent learning. The guarantee of equilibrium convergence is a desirable feature for learning outcomes, in that it satisfies stability and the (admittedly weak) notion of unilateral optimality. Indeed, learning algorithms such as *Nash Q-learning* (J. Hu and Wellman 2003) and *Hidden Markovian Play* (Chen et al. 2015) have been developed expressly for this purpose.

The example of HFT highlights some prominent features that require implementation before these algorithms can be operationalized effectively. Some are well-addressed within the existing literature; others are only partially-addressed or entirely ignored. Of course, any persisting lapses will become more prominent as artificial intelligence (AI) becomes more widespread (therefore, more likely to interact with other AIs) and more relied upon (therefore, required to solve problems that even humans have trouble with – consider the prominent role that anthropomorphism has played in human history).

2 The Desiderata

These features can be divided into those pertaining to the information that algorithms have access to, and those pertaining to the optimality conditions that should be achieved.

Information I: Semi-coupledness. For an algorithm to be uncoupled in a class of games, it must apply to every possible specification of the opponents' preferences. This assumption, widely adopted in the learning literature, is useful in that any algorithm satisfying it will function without any information about others' preferences beyond their representability. It is certainly necessary, in the HFT example, to allow for a wide range of opponent preferences.

[1] In practice, of course, most learning algorithms are not explicitly stated in terms of a prior belief.

However, while this assumption is valid in spirit, it may be excessively strict in practice. There are myriad situations in which additional, credible information may be brought to bear, including structural restrictions on preferences and/or the presence of underlying cognitive limitations. Indeed, it has been conjectured that some of the strict impossibility results on the existence of learning algorithms (Foster and Young 2001; Hart and Mas-Colell 2003)) rely on the unreasonable strength of the uncoupledness assumption.

Information II: Environmental agnosticism. Often, learning algorithms must not only function well regardless of the number of intelligent/unintelligent opponents, but also they must do so without knowing in advance the number and intelligence of said opponents. The importance of this is immediate in the HFT example, although it is worth noting that in many designed markets, the presence of intelligent opponents may be easily trackable. Dekel et al. (2004) considers the difficulty of learning when nature's strategy is unknown. Foster and Young (2006) consider radical uncoupledness in learning algorithms, a condition which requires that (like uncoupledness) one's algorithm not depend on opponent's preferences and (unlike uncoupledness) one's algorithm not depend on opponent's actions. This has the advantage of ensuring environmental agnosticism, but Foster and Young (2006) only prove equilibrium convergence in finite two-player games. Fortunately, as the discussion in section 3 on aggregated actions will suggest, there are weaker sufficient conditions.

Optimality I: Incentive compatibility. The crucial new feature to consider here is incentive compatibility: if all opponents adopt the given learning algorithm, is it optimal to adopt it in response? In HFT, traders have overriding incentives to use the best algorithm available in their environment, without regard for its convergence properties or social inefficiencies. Deviations from existent learning algorithms usually take the form of strategic teaching, wherein agents adjust their behavior to misguide their opponents. This incentive is explored theoretically in Schipper (2015), Israeli (1999), Duersch et al. (2012) and observed experimentally in Hyndman et al. (2012) and Chong et al. (2006). The insuffiency of traditional approaches to learning is highlighted in particular by Schipper (2015), who demonstrates how the assumption that uncoupled learning converge to Nash equilibrium necessarily contradicts incentive compatibility. Finally, note that there differing degrees of incentive compatibility: some market designers may be able to restrict agents to a particular class of strategies, in which case only incentive compatible revelation is required.

Optimality II: Pareto optimality. Beyond unilateral optimality, we may seek learning algorithms that embody some notion of social efficiency. This is not guaranteed: consider the flash crash of 2010, which involved a brief but substantial loss of value on several financial exchanges and was driven by naive responses of HFT algorithms. However, there is reason to be optimistic. Most existing learning algorithms converge to one-shot Nash equilibrium (by design), but it is unclear why one-shot Nash equilibrium should be the standard. While it is necessary that rational learning converge to a Nash equilibrium, the equilibria of relevance are those of the repeated game (Kalai and Lehrer 1993). Consider the repeated prisoner's dilemma: one might expect rational agents to avoid a poor outcome by convincing one another that they deeply care for the other's well-being. Generally, one would hope to design algorithms that converge to some pareto optimal outcome of the repeated game.

To summarize, we seek an incentive-compatible, environmentally-agnostic, semi-coupled, pareto optimal learning algorithm. Satisfying these features would be the crucial necessary step towards employing inference that is robust to the presence of intellectually-similar agents. The importance of such developments to the broader field of AI is clear, but there are two particular applications to economics worth highlighting. The first considers statistical inference about variables of economic relevance, as undertaken within the private and public sectors. When this information is used to inform economically-relevant decisions, we encounter an interactive learning problem. It may be worthwhile to consider whether particular estimation techniques can predict effectively when their usage is widespread.

The second application, as exemplified by HFT, is market design. Here, learning algorithms act as surrogates on behalf of human market participants. This is a form of market design with minimal underlying assumptions on preferences and game structure, which furthermore does not require the designer to compute a constrained optimum (a task that spawned an entire subfield). By commiting agents to behavior in an incentive-compatible way, an appropriate class of learning algorithms could guarantee reasonably good results in a variety of markets – markets that humans currently participate in directly, possibly at great expense and subject to humanity's biological limitations.

The remainder of this proposal will discuss two directions in which this research can proceed, one which we regard as more obvious, more ambitious, but less promising. To aid the discussion, we define our context more formally.

3 Formalizing the Problem

For now, the setting of interest will be a collection Γ^R of repeated games with imperfect monitoring. A learning algorithm (as defined in this section) will have to satisfy optimality criteria (as proposed in sections 4 and 5) for any realized game $\gamma^R \in \Gamma^R$. In particular, the algorithm makes no reference to the likelihood with which a particular game is realized[2]. Instead, Γ^R should be thought of as the domain over which our hypothetical learning algorithm is well-functioning. The collection is generally not the set of all games, and may be subject to whatever restrictions that the designer sees fit.

- Let there be a set Ω of states, endowed with some topology.
- Let Γ be some collection of normal-form games.
  - Let $\mathcal{I}$ be the collection of agent sets associated with some $\gamma \in \Gamma$.
  - Let $\mathcal{A}$ be the collection of action profile sets associated with some $\gamma \in \Gamma$.
- Let $\tau : \mathcal{A} \to \Omega$ prescribe the state associated with any given action profile.
- Let $u_i : \Omega \to \mathbb{R}$ prescribe agent i's utility as a function of the realized state.
  - The utility function for a given game is $u_i \circ \tau$; that is, agent i's utility as a function of the action profile.

From a given $\gamma \in \Gamma$ with agent set I and action profile set A, construct a discrete infinitely repeated game γ^R.

- Define a strategy for agent i as

  $$s_i : \bigcup_{t=0}^{\infty} \left(\prod_{\ell=0}^{t} A_i \times \prod_{\ell=0}^{t} \Omega \right) \to A_i$$

  Accordingly, each agent observes only the state history and their own actions. Note that the structure of a strategy is identical across all repeated games Γ^R.

- Define i's utility $v_i^\delta : S \to \mathbb{R}$ from strategy profile s as the time-discounted sum of stage game utilities for some discount factor $\delta \in (0, 1)$. This may be expressed recursively as:

  $$v_i^\delta(s) = u_i \circ \tau \circ s^\emptyset + \delta v_i^\delta(s(s_i^\emptyset, \tau \circ s^\emptyset, \cdot))$$

  where $s^\emptyset := s(\emptyset)$ is the initial action profile associated with s.

[2] Should it be known *a priori* that games are realized according to some probability distribution μ, the appropriate collection of games would be a singleton, consisting of the incomplete information game with a common prior μ over the game structure. There is another variation, in which agents have prior beliefs μ_i that they would like to see incorporated into the learning algorithm. We might regard this as a retracted learning problem that includes the selection of μ_i given some available information. If (μ_i, μ_{-i}) are "credible", then they should be consistent with a learning algorithm in the retracted learning problem.

- Define a learning algorithm as a function f that takes in preferences $u_i : \Omega \to \mathbb{R}$ and puts out a strategy s_i of the repeated game. For example, we could think of f^∞ as embodying some subjective prior belief over histories, and define $f(u_i)$ as the u_i-optimal play under Bayesian updating.

- For a given game γ^R, define a strategy profile s where $s_i = f^\infty(u_i)$ for all $i \in I$. Later on, we will refer to action and state histories. Construct these sequences by defining $a^s_{i0} = s^\emptyset_i$, $\omega^s_0 = \tau(s^\emptyset)$ and

$$a^s_{it} = s_i(a^s_0, \dots, a^s_{t-1}, \omega^s_0, \dots, \omega^s_{t-1})$$

$$\omega^s_t = \tau \circ s(a^s_0, \dots, a^s_{t-1}, \omega^s_0, \dots, \omega^s_{t-1})$$

  In the presence of mixed strategies, these sequences will be stochastic.

This completes the notation for this proposal. The remaining sections will discuss optimality conditions (primarily, incentive compatibility). Before proceeding, however, we should consider how this framework interacts with our two informational desiderata.

- **Semi-coupledness.** For player i, the learning algorithm f and the strategies it returns depend only on the state history, i's action history, and i's preferences. There is no direct reference to opponents' preferences. By varying the set of games Γ and (through τ) the information that states convey about opponent actions, we can adjust the strength of the semi-coupledness assumption. For example, a typical uncoupled learning algorithm would (a) set Γ to be the space of all n-player normal-form games with action profiles $A = \prod_{j=1}^n A_j$ and (b) set $\Omega = A_{-j}$ so that opponents' actions are fully observed. One plausible weakening would use Γ' as the space of games, where Γ' is the set of $\gamma \in \Gamma$ where every player j satisfies u_i continuous over Ω.

- **Enviromental-agnosticism.** Suppose that we could aggregate opponent's per-period actions in such a way that preferences depended only on the aggregate. This aggregate action would be observed by the learning algorithm as a state $\omega \in \Omega$, along with the individual's present action. If we let $\mathcal{S}$ denote the set of all conceivable aggregated behavior (under any desired preference restrictions), then we might hope for a rich-enough set $\mathcal{I}'$ of two-player agent sets where for any behavior in $\mathcal{S}$ is mimicked by some agent j in an appropriate two-player game. Here, effective learning in the given game would be equivalent to effective learning in the two player game against j, and we could expand $\mathcal{I}$ to include agent sets of arbitrary cardinality and composition.

In this environment, environmental agnosticism is achieved. But the environment (which may be necessary as well as sufficient) is certainly restrictive. It forces the learning algorithm f to perform well in a larger space of possible games (those corresponding to agent sets $\mathcal{I}'$). It also forces f to be flexible enough to exhibit the full range of conceivable aggregate phenomena. These two requirements are complementary in their restrictiveness: a larger space of relevant games will induce a larger range of conceivable phenomena, and more phenomena will require a richer collection $\mathcal{I}'$.

4 The First Approach: Learning Equilibrium

The first approach is to pursue an equilibrium of learning algorithms, a profile of learning algorithms where some notion of incentive compatibility holds for all agents. This may be thought of as an equilibrium in the one-shot game where learning methods (or equivalently, subjective priors) are chosen. More formally, one possible formulation of our objective is as follows. As discussed in the previous section, the following conditions must hold for any $\gamma \in \Gamma$ (and every $i \in I$).

- **Convergence.** The limits $a^s_{i\infty} = \lim_{t\to\infty} a^s_{it}$ and $\omega^s_\infty = \lim_{t\to\infty} \omega^s_t$ exist[3]. If ω^s_∞ has a non-singleton support, then agent i is indifferent between the states realized with positive probability.

- **No-average-regret.** The convergent state ω^s_∞ is payoff-optimal in the set of feasible states for agent i given opponent strategies s_{-i}. Recall that, by the previous bullet, $u_i\left(\omega^s_\infty\right)$ is deterministic. As such[4],

$$u_i\left(\omega^s_\infty\right) = \max_{s'_i} \mathbb{E}\left[\lim_{T\to\infty} \frac{\sum_{t=0}^{T} u_i\left(\omega_t^{(s'_i, s_{-i})}\right)}{T}\right]$$

  This embodies two desiderata. First, agent i will act optimally at the convergent state. Second, agent i will optimally manipulate others to achieve its desired convergent state.

Consider an ideal learning algorithm f^∞ such that $s_i = f^\infty(u_i)$ satisfies these two properties, evaluated under u_i, when $s_{-i} = f^\infty(u_{-i})$. We will seek a sequence of learning algorithms f^k such that as $k \to \infty$, f^k will approximate the aformentioned properties of f^∞.

[3] It is possible that this notion of convergence is too restrictive, in that it rules out convergence to repeated game equilibria with non-stationary outcomes. It will be kept, however, until proven untenable.

[4] Of course, no-average-regret is not always an appropriate formalization of incentive compatibility. Under frequent entry/exit of participating agents, the convergent state may never be reached or may be reached only for a short time. Situations where initial losses preclude future wins (e.g. the agent defaults) would be problematic if our setting were not restricted to repeated games (instead of dynamic games).

It is informative to illustrate this incentive compatibility condition in the abstract. Suppose that f^∞ operates on a collection Γ of two-player games, where Γ is symmetric in the sense that the row player has the same set of possible preferences as the column player. Let a, b, h be functions where a, b represent strategies such that $a = f^\infty(u_1), b = f^\infty(u_2)$ and h_1 represents the payoff evaluation (i.e. no-average regret) as applied to a particular game. We require

$$a = \arg\max_{\forall c} h_1(c, b)$$

This must hold for a variety of functions b. Let $\mathcal{F}$ denote the set of all possible opponent strategies under learning algorithm f^∞. We can write these equations more concisely as

$$a = \arg\max_{\forall c} h_1(c, \mathcal{F}) \tag{1}$$

If we replace references to f^∞ with references to f^k and rewrite equation 1 accordingly, then we would require that equation 1 be met with arbitrarily small error as $k \to \infty$.

While learning equilibrium captures incentive compatibility in the most direct way, it is not clear whether equilibrium learning algorithms are likely to exist. There are several reasons to be concerned.

1. First, a learning algorithm is a complex object, a mapping from functions to other (also complex) functions. These are far more complex, for instance, than the mappings from/into Euclidean space that typically comprise the strategy space of Bayesian games, a type of game for which proving equilibrium existence is also difficult.

2. Second, it is not obvious how an equilibrium learning algorithm should deal with conflicting interests. On one hand, a useful algorithm will have to account for a variety of games and therefore be flexible in its responses. However, given such flexibility, we have an analog to the Schipper (2015) argument: an opponent can always mimic some other type in order to shift the convergent outcome in their favor. Moreover, insofar as accurate prediction is equivalent to optimal behavior, the impossibility result of Foster and Young (2001) suggests that there exist games (in particular, zero-sum games) for which learning equilibrium may never be satisfied.

3. Third, even if f^∞ exists, there may not exist a sequence f^k that approximates learning equilibrium for high k. This is prohibitive when the index k captures a computational limitation that cannot be bypassed, such as a finite memory.

Despite these difficulties, notions of learning equilibrium have arisen within the computer science literature on learning in games, with due consideration of the existence problem.

Shoham et al. (2007) discuss learning equilibrium as one of five agendas for multi-agent learning research, and Monderer and Tennenholtz (2007) elaborate with a fuller exposition of progress in that area. Brafman and Tennenholtz (2004) prove existence in repeated games where opponent preferences are observed and demonstrate that existence is not guaranteed in an uncoupled setting. Brafman and Tennenholtz (2006) prove existence in symmetric games. Ashlagi et al. (2006b) prove existence in an auction setting where only winning bids are observed. In addition, some work has attempted to construct learning equilibrium under weaker unilateral optimality conditions, such as minimax strategies (Hyafil and Boutilier 2004) and safety level equilibrium (Ashlagi et al. 2006a).

It is likely that any non-existence of learning equilibrium is exacerbated by a particular modelling assumption: the absence of computational costs. In the presence of heterogeneous computing costs among agents, one might expect each agent to resign itself to being manipulated by their cognitive superiors who can afford higher-intensity algorithms. It is unclear, however, to what extent the particular specification of computational cost (as well as the payoffs of the agents) will affect the equilibrium learning algorithm.

5 The Second Approach: Learning curb

The second approach takes a lesson from the discussion of computational costs. It attempts to translate competition in the learning strategy space into linear competition in the computing power brought to bear on a specified algorithm. To do this, rather than consider learning equilibrium, we attempt to describe an appropriate curb set: a set that is "closed under rational behavior". (The term was defined by Basu and Weibull (1991) as a set of strategy profiles that includes *all* of its own best replies. Lacking a better term, for the purposes of this proposal we redefine a curb as a set of strategy profile that includes *at least one* best reply to each of its consituents.) The nature of this curb set is very particular. We index strategies by some k that represents sophistication, and require that, given that opponents use a given algorithm at some level of sophistication k_{-i}, we can get arbitrarily close to a best response by using the same algorithm at a sufficiently high k_i.

This curb set is not a learning equilibrium in itself, and may not include the learning equilibrium even if it exists. Why, then, is the emphasis learning curb warranted at all? After all, it does not resolve the indeterminacy problem, where agents are either forced into suboptimal algorithms or constantly increasing their own sophistication in response to corresponding increases by their opponents. To see the value in restricting learning to a self-consistent subset, note that the indeterminacy problem arises only when there is some conflicting interest between participating agents. Consider that our social interest in the outcomes of games of conflict is often quite limited. In the example of HFT, regulators are not concerned about which particular bank captures the surplus in a given transaction. Instead, they are concerned with how uncontrolled competition in these areas of conflict might spill over to destabilize the market in areas of common interest. The primary goal of a learning algorithm designer should be to ensure that, where common incentives do exist, they are attained. Restricting the frame of competition will not resolve unresolvable conflicts, but it may permit stronger efficiency guarantees everywhere else.

To formalize our notion of learning curb, recall the optimal learning algorithm f^∞ of the previous section. As before, consider a sequence of learning algorithms f^k that we intuit as approximating f^∞. Previously, f^k were intended to satisfy the following condition: if all agents adopt f^k, then as $k \to \infty$, the convergence and no-average-regret conditions for all players are realized. Now, f^k is constructed as follows: when all other agents $-i$ adopt $f^{k_{-i}}$ for some finite k_{-i}, as $k_i \to \infty$ the no-average-regret condition for player i is realized. Further, as $k_i, k_{-i} \to \infty$, the convergence condition is realized.

I will illustrate the incentive compatibility condition of $f^1, f^2, ..., f^\infty$ under the same conditions as in the previous section. Here, let $\mathcal{F}$ denote the set of all possible opponent strategies under learning algorithms $f^1, f^2, ..., f^\infty$. Let $\mathcal{F}_i$ denote agent i's possible strategies, namely $\{f^1(u_i), f^2(u_i), ..., f^\infty(u_i)\}$. Let b, h be functions where b represents a strategy such that $b \in \mathcal{F}$ for some k and h_1 represents the payoff evaluation (i.e. no-average regret) as applied to a particular game. We require

$$\exists a \in \mathcal{F}_i \text{ such that } a = \arg\max_{\forall c} h_1(c, b)$$

This must hold for a variety of functions b. If we assume that there is always a unique optimal strategy, we have

$$\mathcal{F}_1 \supseteq \arg\max_{\forall c} h_1(c, \mathcal{F}) \tag{2}$$

While the set $\mathcal{F}$ differs from that of section 4, it is still instructive to compare equations 1 and 2. The former requires a single function to be optimal across a larger set of functions. The latter only requires the optimum to fall within a specified set, a set which can also include functions that are never optimal.

6 Implementation

For any given restriction on preferences, there may be several learning algorithms f, g with that satisfy equations 1 and/or 2. It may be possible to approach the construction of a learning algorithm by first specifying a set of $\mathcal{F}$ behaviors that we wish to capture, and then asking which is the smallest superset $\mathcal{G} \supseteq \mathcal{F}$ that corresponds to an incentive compatible learning algorithm (under some restrictions on $\{h_i\}_{\forall i}$). For example, if we wish to capture all behavior that is linear in the history, is it enough to require that the learning algorithm deal well with linear behavior? Or does it also have to work for, say, all polynomial behavior?

Indeed, there are practical reasons to focus from the beginning on a particular class $\mathcal{F}$ of behavior. There are a number of function classes for which effective statistical prediction tools already exist. If the nature of the learning algorithm and opponent preferences are such that we can guarantee behavior that is linear, or polynomial of degree d, or continuous in recent history, then we can respectively apply ordinary least squares, polynomial estimators, and Taylor approximations to learn with no-average-regret. The relevant question is whether our optimal behavior after applying those estimators will fall into the original class $\mathcal{F}$.

This question is somewhat evocative, and the reasoning resembles a fixed point argument. That intuition can be formalized. Recall equation 1, repeated below for convenience.

$$a = \arg\max_{\forall c} h_1(c, \mathcal{F})$$

This is stated for agent 1, but given any agent i there must exist a^i satisfying this equation under h_i. If we let $\mathcal{H}$ be the set of all possible h_i, then we have the following necessary condition for incentive compatibility.

$$\mathcal{F} = \arg\max_{\forall c} \mathcal{H}(c, \mathcal{F}) \tag{3}$$

This representation suggests that finding an incentive compatible learning algorithm requires solving something analogous to a functional equation, except where the solution is a set of functions rather than a single function. Viewed differently, it is a fixed point of a transformation on sets; namely, the transformation $\arg\max_{\forall c} \mathcal{H}(c, \cdot)$.

There are at least two more properties that may be sought in $\mathcal{F}$. The first is the condition described in section 3, which allows aggregation of multiple players into a single opponent. The second is that belief formation as a function of history is somehow related to one's optimal response as a function of beliefs. For example, beliefs and optimal responses might both be assumed linear in their respective arguments; this would allow for a more sensible interpretation of the requirement that $\mathcal{F}$ consist of linear functions.

Finally, there is the question of how to implement the approximation index. On the one hand, it could prescribe how much information is used, such as only recording the last k periods of history (Hurkens 1995; Powers and Shoham 2005) or only evaluating payoffs up to k periods in the future. On the other hand, the approximation index could prescribe a (limited) level of sophistication at which other agents are assumed to operate. This could either be explicitly rooted in bounded rationality models (Lipman 1991; Camerer et al. 2004; T.-W. Hu 2014), or be stated directly in terms of observed behavior, like assuming opponent's strategies are polynomials (in recent history) of order k. Of course, there is no reason to rule out combinations, where both the information used and the complexity of inference is increasing in k. Indeed, that may be the best approach, in that it allows the designer to approximate rather than solve the various facets of this difficult problem.

References

Ashlagi, I., Monderer, D., & Tennenholtz, M. (2006a). Resource selection games with unknown number of players. In *Proceedings of the fifth international joint conference on autonomous agents and multiagent systems* (pp. 819–825). AAMAS '06. Hakodate, Japan: ACM.

Ashlagi, I., Monderer, D., & Tennenholtz, M. (2006b). Robust learning equilibrium. In *Proceedings of the 22nd conference on uncertainty in artificial intelligence*. UAI '06. Arlington, Virginia, United States: AUAI Press.

Basu, K. & Weibull, J. W. (1991). Strategy subsets closed under rational behavior. *Economics Letters*, *36*(2), 141–146.

Brafman, R. I. & Tennenholtz, M. (2004). Efficient learning equilibrium. *Artificial Intelligence*, *159*(1), 27–47.

Brafman, R. I. & Tennenholtz, M. (2006). Optimal efficient learning equilibrium: imperfect monitoring in symmetric games. In *Proceedings of the 21st national conference on artificial intelligence*. AAAI-06.

Brown, G. W. (1951). Iterative solution of games by fictitious play. In T. C. Koopmans (Ed.), *Activity analysis of production and allocation* (Chap. 24, pp. 374–376). John Wiley & Sons, Inc.

Camerer, C. F., Ho, T.-H., & Chong, J.-K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, *119*(3), 861–898.

Chen, W., Chen, Y., & Levine, D. K. (2015). A unifying learning framework for building artificial game-playing agents. *Annals of Mathematics and Artificial Intelligence*, *73*(3), 335–358.

Chong, J.-K., Camerer, C. F., & Ho, T. H. (2006). A learning-based model of repeated games with incomplete information. *Games and Economic Behavior*, *55*(2), 340–371.

Dekel, E., Fudenberg, D., & Levine, D. K. (2004). Learning to play bayesian games. *Games and Economic Behavior*, *46*(2), 282–303.

Duersch, P., Oechssler, J., & Schipper, B. C. (2012). Unbeatable imitation. *Games and Economic Behavior*, *76*(1), 88–96.

Foster, D. P. & Vohra, R. V. (1997). Calibrated learning and correlated equilibrium. *Games and Economic Behavior*, *21*(1), 40–55.

Foster, D. P. & Young, H. P. (2001). On the impossibility of predicting the behavior of rational agents. *Proceedings of the National Academy of Sciences*, *98*(22), 12848–12853.

Foster, D. P. & Young, H. P. (2006). Regret testing: learning to play nash equilibrium without knowing you have an opponent. *Theoretical Economics*, *1*(3), 341–367.

Fudenberg, D. & Levine, D. K. (1993). Self-confirming equilibrium. *Econometrica*, *61*(3), 523–545.

Fudenberg, D. & Levine, D. K. (1998, December). *The Theory of Learning in Games*. MIT Press Books. The MIT Press.

Hart, S. & Mas-Colell, A. (2003). Uncoupled dynamics do not lead to nash equilibrium. *The American Economic Review*, *93*(5), 1830–1836.

Hu, J. & Wellman, M. P. (2003). Nash q-learning for general-sum stochastic games. *Journal of Machine Learning Research*, *4*, 1039–1069.

Hu, T.-W. (2014). Unpredictability of complex (pure) strategies. *Games and Economic Behavior*, *88*, 1–15.

Hurkens, S. (1995). Learning by forgetful players. *Games and Economic Behavior*, *11*(2), 304–329.

Hyafil, N. & Boutilier, C. (2004). Regret minimizing equilibria and mechanisms for games with strict type uncertainty. In *Proceedings of the 20th conference on uncertainty in artificial intelligence* (pp. 268–277). UAI '04. Banff, Canada: AUAI Press.

Hyndman, K., Ozbay, E. Y., Schotter, A., & Ehrblatt, W. Z. (2012). Convergence: an experimental study of teaching and learning in repeated games. *Journal of the European Economic Association*, *10*(3), 573–604.

Israeli, E. (1999). Sowing doubt optimally in two-person repeated games. *Games and Economic Behavior*, *28*(2), 203–216.

Kalai, E. & Lehrer, E. (1993). Rational learning leads to nash equilibrium. *Econometrica*, *61*(5), 1019–1045.

Lipman, B. L. (1991). How to decide how to decide how to...: modeling limited rationality. *Econometrica*, *59*(4), 1105–1125.

Monderer, D. & Tennenholtz, M. (2007). Learning equilibrium as a generalization of learning to optimize. *Artificial Intelligence*, *171*(7), 448–452.

Powers, R. & Shoham, Y. (2005). Learning against opponents with bounded memory. In *Proceedings of the 19th international joint conference on artificial intelligence* (pp. 817–822). IJCAI'05. Edinburgh, Scotland: Morgan Kaufmann Publishers Inc.

Schipper, B. C. (2015, April). *Strategic teaching and learning in games.*

Shoham, Y., Powers, R., & Grenager, T. (2007). If multi-agent learning is the answer, what is the question? *Artificial Intelligence*, *171*(7), 365–377.

Young, H. P. (2004, December). *Strategic Learning and its Limits.* OUP Catalogue. Oxford University Press.