Noname manuscript No.
(will be inserted by the editor)

Agent-Based Modeling of a Non-Tâtonnement Process for the Scarf Economy

The Role of Learning

Shu-Heng Chen · Bin-Tzong Chie · Ying-Fang Kao · Ragupathy Venkatachalam

Received: date / Accepted: date

Abstract In this paper, we propose a meta-learning model to hierarchically integrate individual learning and social learning schemes. This meta-learning model is incorporated into an agent-based model to show that Herbert Scarf's famous counterexample on Walrasian stability can become stable in some cases under a non-tâtonnement process when both learning schemes are involved, a result previously obtained by Herbert Gintis. However, we find that the stability of the competitive equilibrium depends on *how individuals learn* - whether they are innovators (individual learners) or imitators (social learners), and their switching frequency (mobility) between the two. We show that this endogenous behavior, apart from the initial population of innovators, is mainly determined by the agents' intensity of choice. This study grounds the Walrasian competitive equilibrium based on the view of a balanced resource allocation between exploitation and exploration. This balance, achieved through a meta-learning model, is shown to be underpinned by a behavioral/psychological characteristic.

Shu-Heng Chen, Ying-Fang Kao
AI-ECON Research Center,
Department of Economics,
National Chengchi University
Taipei, Taiwan 11605
E-mail: chen.shuheng@gmail.com, seldakao@gmail.com

Bin-Tzong Chie
Department of Industrial Economics,
Tamkang University,
Taipei, Taiwan 25137
E-mail: chie@mail.tku.edu.tw

Ragupathy Venkatachalam
Institute of Management Studies,
Goldsmiths, University of London,
London SE14 6NW, UK
E-mail: rpathy@gmail.com

Keywords Non-tâtonnement process · Co-ordination · Agent-Based modeling · Learning

1 Introduction

Market economies are often seen as exhibiting a 'remarkable degree of coherence' (Arrow, 1974) in terms of coordinating a highly dispersed multitude of actions, performed by very many individuals. How does the economic system achieve this coordination? Developments in the field of general equilibrium theory over the last 140 years have shown the conditions under which such a coherence may be plausible. However, coherence is interpreted in terms of an equilibrium phenomenon in this class of models. The discovery of equilibrium prices that balance the demands and supplies of various actors in the aggregate, across all markets, is seen as being mediated through a fictional, centralized authority. This authority, known as the *Walrasian auctioneer*, supposedly achieves this discovery through a process of trial and error (tâtonnement).

In this framework, trading happens, if at all it happens, only in equilibrium. Despite its advantages, tâtonnement as a tool to find a solution needs to be distinguished from both the plausibility and reality of the process of such an interaction[1]. The tâtonnement process is highly divorced from how agents actually go about achieving coordination by searching and learning to discover prices. Therefore, it becomes necessary to go beyond the tâtonnement process and find plausible descriptions of behavior that respect the limitations that human agents face. [2]

Attempts to weaken the assumption of a centralized auctioneer, allowing for disequilibrium trading, has also been the theme of research on *non-tâtonnement processes* (Hahn and Negishi, 1962; Uzawa, 1960; Fisher, 1983)[3]. To permit disequilibrium trading, however, means that there are many trading protocols that are possible, and developing behaviorally plausible and acceptable trading protocols becomes important. Given such a decentralized, disequilibrium trading process, we can ask whether the economy will steer itself toward the competitive general equilibrium.

To this end, viewing an economy as a decentralized, distributed computing system to cope with the problem of coordinating the demand and supply of various agents can be fruitful. In this framework, no one person has complete knowledge of the entire system. However, agents interacting amongst themselves could collectively solve this dynamic problem of price discovery by trial

[1] Velupillai (2015) succinctly summarizes the different facets of the problem, viz. the 'existence of a solution, a method of finding it (if 'proved' to exist), the 'reality' of the method considered as a dynamic process and its stability' (Ibid, p. 1556).

[2] This is expressed with admirable clarity in Clower (1975), p.13.

[3] Several important contributions made in the 1970s and 1980s also took up the disequilibrium and out-of-equilibrium dynamics to establish a link between Keynesian and general equilibrium models. On this see, Benassy (1982); Malinvaud (1977).

and error.[4] Agent-based modeling provides a suitable framework to investigate questions of such nature (Tesfatsion, 2006). Albin and Foley (1992) made initial attempts to study decentralized interaction without an auctioneer and more recently a few studies - Gintis (2006, 2007, 2013),Mandel (2012), Mandel et al. (2015)- have advanced this approach by augmenting agents with learning capabilities.

This latest development, referred to as *Walrasian dynamics* by Herbert Gintis, distinguishes itself by showing how the non-tâtonnement processes to the general equilibrium may work even though the tâtonnement processes fail. For instance, Scarf (1960) demonstrated an important counterexample to the stability of the Walrasian general equilibrium under the tâtonnement process; Gintis (2007), nonetheless, has shown that the Walrasian equilibrium as a globally stable stationary state can ensue in the Scarf model by employing agent-based simulations. Gintis (2013) further indicates that neither individual nor social learning alone can ensure price discovery. To converge to the competitive equilibrium (CE), we need both.

> . . . individual learning is insufficient to produce market equilibrium. The imitation process is much more powerful than individual learning, but the two in combination are quite powerful even in the case of many goods. (Ibid, p. 123)

We consider the above quotation to be quite intriguing and deserving further elaboration. First of all, individual learning (sometimes also recognized as exploitation or innovation) and social learning (exploration or imitation) are two different types of learning from experiences. It is interesting to see whether these two learning schemes can result in different outcomes (Vriend, 2000). However, few studies have questioned the combined effect of the two and their possible synergies. Second, while Gintis did notice the significance of their combined use, he did not further distinguish different forms of such combinations. He considers *sequential* combination, i.e., in two stages, where agents sequentially apply social learning (imitation) and individual learning (mutation). This form is rather familiar in evolutionary game theory (Samuelson, 1998). An alternative is a *hierarchical* combination. In this form, at each point in time, each agent can either decide to innovate (individual learning) or imitate (social learning), but not both. This latter form is frequently seen in evolutionary computation (Koza, 1992), but was not studied by Gintis.

By and large, sequential combination is biologically motivated, which interprets mutation as an ‘error’ added to the self-reproduction process of chromosomes. Being an error, this kind of biological mutation normally leads to a disaster. Therefore, when ‘blindly’ applied to social sciences, it does not help answer the basic question: what motivates the agent to randomly change the copied strategy (price)? The motivation is normally not part of the model. On the other hand, hierarchical combination has no room for this ‘natural

[4] At a systemic level, it has been shown that the computational complexity of a decentralized system of interacting agents is lower than that of the centralized system that is based on a Walrasian auctioneer (Axtell, 2005).

mistake': imitation and innovation are two separate choices. The motivation behind this choice problem is a matter of personal preference and may depend on the agent's experience.

In this paper, we introduce a meta-learning approach for agents to decide when to learn from others' experiences (imitation or social learning) and when to learn from his/her own experience (innovation or individual learning). The meta-learning model is concretized by reinforcement learning. This formulation is similar to the *heuristic switching models* popularly used in the agent-based computational economics literature (Hommes, 2011; Anufriev and Hommes, 2012; Hommes & Zeppini, 2014). Needless to say, the value of our proposal does not lie in this specific form of meta-learning; instead, it lies in its extendibility to other equally plausible behavioral heuristics.

Through the meta-learning model, hierarchical combination not only allows agents to endogenously determine whether to imitate or innovate, but also endogenously generate the population size of imitators and innovators over time. This degree of autonomy and heterogeneity helps us to gain insights into the market mechanism in terms of constantly searching for a balance between exploitation and exploration as a way of knowledge production and distribution. This balance is not accomplished by a central planner in the sense of the Walrasian auctioneer, but by a 'man on the spot' in the spirit of Hayek (1945).

In this paper, we revisit Gintis' Walrasian dynamics in an agent-based Scarf economy by engaging in a systematic and thorough study of the combined use of individual learning and social learning. We investigate the price dynamics with different parameter values of the proposed meta-learning model. Like Gintis, we apply the *simplicity principle* (Axelrod, 1997) and only consider the simplest form of individual and social learning, namely, mutation and imitation.

Our major findings can be summarized as follows. As in Gintis (2007, 2013), we show that neither individual learning nor social learning alone can always help agents coordinate their prices and lead the economy to the competitive equilibrium. Moreover, we also show that an economy composed of an exogenously fixed fraction of imitators and innovators, without switching between the two groups, is also insufficient to ensure price coordination. However, if we allow agents to decide when to imitate and innovate through reinforcement meta-learning, then the economy can reach its equilibrium configuration when a parameter known as the *intensity of choice* is properly set. In fact, this parameter endogenously determines the fraction of innovators and imitators, the fraction of mobile (switching) agents, and the resources allocated between exploration (being analogous to social learning) and exploitation (being analogous to individual learning), which in turn determines the price deviation away from the competitive equilibrium.

The rest of the paper is organized as follows: Section 2 introduces the Scarf economy and presents an agent-based version of the same along with the bilateral trading protocols.[5] Section 3 presents the learning mechanisms used by the agents in this economy. Section 4 outlines the simulation design followed by a presentation of the simulation results in Section 5 and concluding remarks in Section 6.

2 The Scarf Economy and the Non-Tâtonnement Process

Scarf (1960) showed that an economy may not always converge to the Walrasian general equilibrium under the tâtonnement process, even when the equilibrium is unique. This raises concerns about the stability of the general equilibrium. In an attempt to provide a dynamic and evolutionary foundation for a general equilibrium, Gintis (2007) used an agent-based model to demonstrate that when agents hold subjective expectations on prices (referred to as private prices)[6] and imitate other agents who have been successful, prices converge to the unique general equilibrium of the Scarf-like economy. As in Gintis, we start with the Scarf economy because of its structure that is readily amenable to investigation and the popularity that it has enjoyed.

2.1 An Agent-Based Model of the Scarf Economy

Following Scarf (1960), we consider a pure exchange economy composed of N agents. This economy has three goods, denoted by $j = 1, 2, 3$, and correspondingly, N agents are grouped in to three different 'types', τ_j, $j = 1, 2, 3$, where $\tau_1 \equiv \{1, \dots, N_1\}; \tau_2 \equiv \{N_1 + 1, \dots, N_1 + N_2\}; \tau_3 \equiv \{N_1 + N_2, \dots, N\}$. Agents who belong to τ_j (type-j) are initially endowed with w_j units of good j, and zero unit of the other two goods. Let $\mathbf{W_i}$ be the endowment vector of agent i:

$$\mathbf{W_i} = \begin{cases} (w_1, 0, 0), & i \in \tau_1, \\ (0, w_2, 0), & i \in \tau_2, \\ (0, 0, w_3), & i \in \tau_3. \end{cases} \tag{1}$$

All agents are assumed to have a Leontief type payoff function.

$$U_i(x_1, x_2, x_3) = \begin{cases} \min\{\frac{x_2}{w_2}, \frac{x_3}{w_3}\}, & i \in \tau_1, \\ \min\{\frac{x_1}{w_1}, \frac{x_3}{w_3}\}, & i \in \tau_2, \\ \min\{\frac{x_1}{w_1}, \frac{x_2}{w_2}\}, & i \in \tau_3. \end{cases} \tag{2}$$

[5] In order to examine the role played by learning, we choose a version that is closer to Scarf's original version and augment it with learning. It is therefore different from Gintis (2007) , which has production.

[6] As Gintis (2013, p.119) notes:

> ...This is the fact that in a decentralized market economy out of equilibrium, there is no price vector for the economy at all. The assumption that there is a system of prices that are common knowledge to all participants (we may call these public prices) is reasonable in equilibrium, because all agents can, at least in principle, observe the same prices. However, out of equilibrium there is no single set of prices determined by market exchange. Rather, every agent has a subjective prior concerning prices, based on personal experience, that he uses to make and carry out trading plans.

Given the complementarity feature of this payoff function, we populate an equal number of agents in each type. This ensures that the economy is balanced and that no free goods appear.

We assume that agents have their own subjective expectations of the prices of different goods,

$$\mathbf{P^e_i(t)} = (P^e_{i,1}(t), P^e_{i,2}(t), P^e_{i,3}(t)), \quad i = 1, ..., N. \tag{3}$$

where $P^e_{i,j}(t)$ is agent i's price expectation of good j at time t.[7]

Given a vector of the subjective prices (private prices), the optimal demand vector, $\mathbf{X}^* = \Psi^*(\mathbf{W})$ can be derived by maximizing payoffs with respect to the budget constraint.

$$\mathbf{X^*_i} = (x^*_{i,1}, x^*_{i,2}, x^*_{i,3}) = \begin{cases} (0, \psi^*_i w_2, \psi^*_i w_3), & i \in \tau_1, \\ (\psi^*_i w_1, 0, \psi^*_i w_3), & i \in \tau_2, \\ (\psi^*_i w_1, \psi^*_i w_2, 0), & i \in \tau_3. \end{cases} \tag{4}$$

where the multiplier

$$\psi^*_i = \begin{cases} (P^e_{i,1} w_1)/(\sum_{j=2,3} P^e_{i,j} w_j), & i \in \tau_1, \\ (P^e_{i,2} w_2)/(\sum_{j=1,3} P^e_{i,j} w_j), & i \in \tau_2, \\ (P^e_{i,3} w_3)/(\sum_{j=1,2} P^e_{i,j} w_j), & i \in \tau_3. \end{cases} \tag{5}$$

Note that the prices in the budget constraint are 'private' prices. Rather than restricting the prices within a certain neighborhood (for instance a unit sphere in Scarf (1960)), we follow Anderson et al. (2004) and set one of the prices (P_3) as the numéraire. The Walrasian competitive equilibrium for this system is $P^*_1 = P^*_2 = P^*_3 = 1$, when the endowments for each *type* are equal and symmetric. However, in this model agents may have own price expectations that may be very different from this competitive equilibrium price, and base their consumption decision on their private price expectations $P^e_{i,j}$. To facilitate exchange, we match agents in the model randomly with each other and they are allowed to trade amongst each other if they find it to be beneficial. For this, we need to specify a precise bilateral trading protocol and the procedures concerning how agents dynamically revise their subjective prices.

2.2 Trading protocol

We randomly match a pair of agents, say, i and i'. Let i be the *proposer*, and i' be the *responder*. Agent i will initiate the trade and set the price, and agent i' can accept or decline the offer. We check for the double coincidence of wants, i.e., whether they belong to the same type. If they do, they will be rematched; otherwise, we check whether the agents have a positive amount of endowment in order for them to engage in trade. Let m_i be the commodity

[7] More specifically, t refers to the whole market day t, i.e., the interval $[t, t-1)$.

that i is endowed with $(m_i = 1, 2, 3)$. Agent i, based on his subjective price expectations, proposes an exchange to agent i',

$$x^*_{i,m'_i} = \frac{P^e_{i,m_i} x_{i',m_i}}{P^e_{i,m_{i'}}}. \tag{6}$$

Here, the proposer (i) makes an offer to satisfy the need of the agent i' up to x_{i',m_i} in exchange for his own need $x^*_{i,m_{i'}}$.

Agent i' will evaluate the 'fairness' of the proposal using his private, subjective expectations and will consider the proposal interesting provided that

$$P^e_{i',m_i} x_{i',m_i} \geq P^e_{i',m_{i'}} x^*_{i,m_{i'}}; \tag{7}$$

Otherwise, he will decline the offer. Since the offer is in a form of *take-it-or-leave-it* (no bargaining), this will mark the end of trade.

Agent i' will accept the proposal if the above inequality (7) is satisfied, and if $x_{i',m_i} \leq x^*_{i',m_i}$. This *saturation condition* ensures that he has enough goods to trade with i and meet his demand. However, if only the saturation condition is not satisfied, the proposal will still be accepted, but the trading volume will be adjusted downward to $x_{i,m_{i'}} < x^*_{i,m_{i'}}$. Agents update their (individual) excess demand and as long as they have goods to sell, they can continue to trade with other agents. The agents are rematched many times to ensure that the opportunities to trade are exhausted. Once the bilateral exchange is completed, the economy enters into the consumption stage and the payoff of each agent $U_i(\mathbf{X_i(t)}), i = 1, 2, ..., N$, is determined. Note that $\mathbf{X_i(t)}$, the realized amount after the trading process may not be the same as the planned level $\mathbf{X^*_i(t)}$. This may be due to misperceived private prices and a sequence of 'bad luck', such as running out of search time (number of trials) before making a deal, etc. Based on the difference between $\mathbf{X_i(t)}$ and $\mathbf{X^*_i(t)}$, each agent i in our model will adaptively revise his or her private prices through a process of learning.

3 Learning

In this section, we shall follow Gintis (2007) and consider only the simplest form of individual learning and social learning, since our focus is on the proposed hierarchical combination of individual and social learning. We are interested in knowing the price discovery capability when only one scheme of learning is applied, to be compared with the case when both are used. In this article, we do not consider more deliberated, complex forms of individual learning, such as simulated annealing, artificial neural networks, decision tress, fuzzy inference, and social learning, such as genetic algorithms and genetic programming.[8]. They are certainly important branches to consider along the expansion of this research agenda, but to avoid unnecessary complications

[8] For a more extended list, see Chen, Kao and Ragupathy (2016)

they are not considered at this stage. In fact, the simplicity principle prompts us to wonder: if one can obtain a well-performing result by combining two simple learning rules, then why bother using more complex ones? Consequently, for the simplest form of individual learning, we consider just *mutation*, symbolizing innovation (Brenner, 1998), and, for social learning, we simply consider *imitation* (Rendell et al., 2010).

3.1 Individual Learning

We begin with the mutation-like individual learning. Mutation can be understood as a perturbation to the incumbent strategy; in our case, this applies to the behavior of price expectations, $\mathbf{P_i^e(t)}$ (see also Equation (3)). It has the following canonical form:

$$\mathbf{P_i^e(t+1)} = \mathbf{P_i^e(t)} + \mathbf{\Delta P_i}, \quad i = 1, 2, ..., N. \tag{8}$$

For the perturbation term $\mathbf{\Delta P_i}$, the mechanism that we employ can be thought of as a modified agent-level version of the Walrasian price adjustment equation. Let the optimal and actual consumption bundles be denoted by the vectors $\mathbf{X_i^*(t)}$ and $\mathbf{X_i(t)}$, respectively. Agent i can check his excess supply and excess demand by comparing these two vectors component-wise. Agents then reflect on how well their strategy (private price expectations) has performed in the previous trading round and adjust their strategies based on their own experience of excess demand and supply. We employ a gradient descent approach to characterize individual learning and, in a generic form, it can be written as:

$$\mathbf{P_i^e(t+1)} = \mathbf{P_i^e(t)} + \underbrace{\mathbf{\Delta P_i(X_i(t), X_i^*(t))}}_{\text{gradient descent}}, \quad i = 1, 2, ..., N, \tag{9}$$

and we shall detail its specific operation as follows.

Let $\mathbf{m}_i^y$ and $\mathbf{m}_i^c$ denote the production set and consumption set of agent i. In the Scarf economy, $\mathbf{m}_i^y \cap \mathbf{m}_i^c = \emptyset$ and in this specific 3-good scarf economy, $\mathbf{m}_i^y = \{m_i\}$. At the end of each market period, agent i will review his expectations for all commodities, $P_{i,j}^e(t), \forall j$. For the good that the agent 'produces' (or is endowed with), the price expectations $P_{i,j}^e(t)$ will be adjusted downward if m_i is not completely sold out (i.e., when there is excess supply). Nonetheless, even if m_i has been completely sold out, it does not mean that the original price expectation will be sustained. In fact, under these circumstances, there is still a *probability* that $P_{i,j}^e(t)$ may be adjusted *upward*. This is to allow agent i to explore or experiment regarding whether his produced (endowed) commodity might deserve a better price. However, to ensure that our agents do not over-react despite having zero-inventory, we assume that their tendency to change prices will decline with the passage of time. That is to say, when agents constantly learn from their experiences, they gradually gain confidence in their zero-inventory price expectations. Specifically, the time-decay function applied in our model is exponential, which means that this kind of exploitation disappears quickly with time. For those goods in the vector $\mathbf{X}_i$ that are a part of the consumption set of the agent, i.e., $j \in \mathbf{m}_i^c$, the mechanism would operate exactly in the opposite manner by increasing the price expectations if there were excess demand. The individual learning protocol is summarized below.

Protocol: Individual Learning

1. At the end of the trading day, agent i examines the extent to which his planned demand has been satisfied. Let

$$\mathit{\Delta} x_{i,j}(t) = \begin{cases} x^*_{i,j}(t) - x_{i,j}(t), & \text{if } j \in \mathbf{m}_i^c, \\ 0 - x_{i,j}(t), & \text{if } j \in \mathbf{m}_i^y \end{cases} \tag{10}$$

2. The subjective prices $P^e_{i,j}$ of all three goods will be adjusted depending on the absolute value $\mid \mathit{\Delta} x_{i,j}(t) \mid$.
3. If $\mid \mathit{\Delta} x_{i,j}(t) \mid > 0$ (i.e., $\mid \mathit{\Delta} x_{i,j}(t) \mid \neq 0$),

$$P^e_{i,j}(t+1) = \begin{cases} (1 + \alpha(\mid \mathit{\Delta} x_{i,j}(t) \mid))P^e_{i,j}(t), & \text{if } j \in \mathbf{m}_i^c. \\ (1 - \alpha(\mid \mathit{\Delta} x_{i,j}(t) \mid))P^e_{i,j}(t), & \text{if } j \in \mathbf{m}_i^y. \end{cases} \tag{11}$$

   where $\alpha(.)$ is a *hyperbolic tangent function*, given by:

$$\alpha(\mid \mathit{\Delta} x_{i,j}(t) \mid) = \tanh(\varphi \mid \mathit{\Delta} x_{i,j}(t) \mid) = \frac{e^{(\varphi|\mathit{\Delta} x_{i,j}(t)|)} - e^{(-\varphi|\mathit{\Delta} x_{i,j}(t)|)}}{e^{(\varphi|\mathit{\Delta} x_{i,j}(t)|)} + e^{(-\varphi|\mathit{\Delta} x_{i,j}(t)|)}} \tag{12}$$

4. If $\mid \mathit{\Delta} x_{i,j}(t) \mid = 0$,

$$P^e_{i,j}(t+1) = \begin{cases} (1 - \beta(t))P^e_{i,j}(t), & \text{if } j \in \mathbf{m}_i^c. \\ (1 + \beta(t))P^e_{i,j}(t), & \text{if } j \in \mathbf{m}_i^y. \end{cases} \tag{13}$$

   where β is a random variable, and is a function of time.

$$\beta = \theta_1 \exp \frac{-t}{\theta_2}, \quad \theta_1 \sim U[0, 0.1], \tag{14}$$

   where $U[0, 0.1]$ is the uniform distribution between 0 and 0.1, and θ_2 is a time scaling constant.

3.2 Social Learning

The second scheme of learning available to the agent is social learning. This process of acquiring relevant information from other members can be decisive for effective adaptation and evolutionary survival. Consequently, social learning has been a prominent feature in the evolutionary models in biology, economics and game theory (Ellison and Fudenberg, 1993; Apesteguia et al., 2007). Imitation is one of the most commonly invoked, simplest forms of social learning. Players imitate for a variety of reasons and the advantages can be in the form of lower information-gathering costs and information processing costs, and imitation also may act as a coordination device in games (Alós-Ferrer & Schlag, 2009).

In this paper, we adopt a fairly basic version of imitation behavior, where agents exchange their experiences regarding payoffs with other agents with whom they are randomly matched. An agent with a lower payoff can, on observing others, replace his own expectations with those of the agent with a higher payoff. If the agent ends up meeting someone who has performed worse than him, he does not imitate and retains his original price expectations. We present the social learning protocol below.

Protocol: Social Learning

1. At the end of each day, each agent consumes the bundle of goods that he has obtained after trading, and derives pleasure from his consumption $U_i(t)$ $(i = 1, .., N)$.
2. Agents are matched randomly, either with other agents of the same type or with agents who are of different types. This is achieved by randomly picking up a pair of agents (i, i') *without replacement* and they are given a chance to interact.
3. Their payoffs are ranked, and the price expectations are modified as follows:

$$\mathbf{P}_i^e(t+1) = \begin{cases} \mathbf{P}_{i'}^e(t), & \text{if } U_i(t) < U_{i'}(t), \\ \mathbf{P}_i^e(t), & \text{if } U_i(t) > U_{i'}(t), \\ \text{Random}(\mathbf{P}_i^e(t), \mathbf{P}_{i'}^e(t)), & \text{if } U_i(t) = U_{i'}(t). \end{cases} \tag{15}$$

The protocol makes it possible for agents to meet locally and enables them to exchange information. An agent who has performed well can influence someone who hasn't performed as well by modifying their perception of the economy.

3.3 Meta Learning

The novelty of the paper is that we propose a meta-learning model (a stochastic choice framework) to hierarchically integrate individual and social learning schemes. In the literature, the only study which can be related to our approach is Bossan, Jann, and Hammerstein (2015). However, in their model, meta-learning is not applied to individuals but to the whole population; in other words, they used a *mesoscopic* approach to determine the proportion (market fraction) of 'individual learners' and 'social learners', an approach similar to the replicator dynamics in evolutionary game theory. We, on the other hand, used a truly *microscopic approach* to leave each individual agent to choose one of the two learning schemes.[9] In our model, each agent can consciously

[9] It is known that these two approaches can generally lead to different results (Grimm and Railsback, 2005). We, however, will leave this issue to a separate study.

choose between social learning and individual learning on each market day based on the past performance of these schemes. We formulate this choice between different learning schemes as a two-armed bandit problem.

3.3.1 Two-Armed Bandit Problem

In our market environment, the agent repeatedly chooses between two learning schemes, individual learning and social learning. Denote the action space (feasible set of choice) by $\mathit{\Gamma}$, $\mathit{\Gamma} = \{a_{il}, a_{sl}\}$, where a_{il} and a_{sl} refer to the action of individual learning and social learning, respectively. Each action chosen at time t by agent i yields a payoff $\pi(a_{k,t})(k = il, sl)$. This payoff is uncertain, but the agent can observe this payoff ex-post and this information is used to guide future choices. This setting is analogous to the familiar *two-armed bandit* problem. In the literature, reinforcement learning has been taken as a standard behavioral model for this type of choice problem (Arthur, 1993). In this paper, we shall follow this tradition. Our reinforcement learning model is to be introduced in Section 3.3.2.

3.3.2 Reinforcement Learning

Reinforcement learning has been widely investigated both in artificial intelligence (Sutton and Barto, 1998; Wiering & van Otterlo, 2012) and economics (Roth and Erev, 1995; Erev and Roth, 1998). The intuition behind this learning scheme is that better performing choices are reinforced over time and those that lead to unfavorable or negative outcomes are not. In our model, each agent reinforces only two choices, i.e., individual learning and social learning ($\mathit{\Gamma} = \{a_{il}, a_{sl}\}$). In terms of reinforcement learning, the probability of a scheme being chosen depends on the (normalized) propensity accumulated over time. Specifically, the mapping between the propensity and the choice probability is represented by the following Gibbs-Boltzmann distribution:

$$Prob_{i,k}(t+1) = \frac{e^{\lambda \cdot q_{i,k}(t)}}{e^{\lambda \cdot q_{i,a_{il}}(t)} + e^{\lambda \cdot q_{i,a_{sl}}(t)}}, \quad k \in \{a_{il}, a_{sl}\}, \tag{16}$$

where $Prob_{i,k}(t)$ is the choice probability for learning scheme k ($k = a_{il}, a_{sl}$); we index this choice probability by i (the agent) and t (time) considering that different agents may have different experiences with respect to the same learning scheme, and that, even for the same agent, experience may vary with time. The notation $q_{i,k}(t)$ ($k = a_{il}, a_{sl}$) denotes the propensity of the learning scheme k for agent i at time t. Again, it is indexed by t because the propensity is revised from time to time based on the accumulated payoff.

The propensity updating scheme applied here is the one-parameter version of Roth and Erev (1995).[10]

$$q_{i,k}(t+1) = \begin{cases} (1-\phi)q_{i,k}(t) + U_i(t), & \text{if } k \text{ is chosen,} \\ (1-\phi)q_{i,k}(t), & \text{otherwise,} \end{cases} \tag{17}$$

where $U_i(t) \equiv U_i(\mathbf{X_i(t)})$, $k \in \{a_{il}, a_{sl}\}$, and ϕ is the so-called recency parameter, which can be interpreted as a memory-decaying factor.[11] The notation λ denotes the intensity of choice. With higher λs, the agent's choice is less random and is heavily biased toward the better performing behavioral strategy; in other words, the degree of exploration in which the agent engages is reduced. In the limit as $\lambda \to \infty$, the agent's choice degenerates to the greedy algorithm which is only interested in the "optimal choice" that is conditional on the most recent updated experience; in this case, the agent no longer explores.

3.3.3 Reference Points

We further augment the standard reinforcement learning model with a reference-point mechanism to decide when Equation (16) will be triggered. Reference dependence in decision making has been made popular by *prospect theory* (Kahneman and Tversky, 1979), where gains and losses are defined relative to a reference point. This draws attention to the notion of *position* concerning the stimuli and the role it plays in cognitive coding that describes the agent's perception of the stimuli.

The meta-learning behavior described by Equation (16) *alone* essentially assumes that agents examine their learning scheme in each period t, with the same frequency that they examine their price expectations. With the augmented version of the meta-learning model, agents are assumed to question the appropriateness of their 'incumbent' learning scheme *only* when their payoffs fall short of the reference point (along the lines of Erev and Rapoport (1998), p.152-153). Let $U_i(t)(\equiv U_i(\mathbf{X_i(t)}))$ be the payoff of an agent i at time t. Let his reference point at time t be $R_i(t)$. The agent will consider a scheme switch *only* when his realized payoff $U_i(t)$ is lower than his reference point $R_i(t)$. In this narrow sense, our agents can be considered to be Simonian *satisficing* agents: they would consider switching the learning scheme only when they are not *satisfied* ($U_i(t) < R_i(t)$); otherwise, they will stick to their current learning scheme.

[10] We certainly can consider more generalized propensity updating dynamics with three parameters as proposed by Camerer and Ho (1999), but that can complicate our analysis at this initial stage. Hence, we plan to start with this 'minimal' model.

[11] By following Arthur (1993), a normalization scheme is also applied to normalize the propensities $q_{i,k}(t+1)$ as follows:

$$q_{i,k}(t+1) \leftarrow \frac{q_{i,k}(t+1)}{q_{i,a_{il}}(t+1) + q_{i,a_{sl}}(t+1)}. \tag{18}$$

The reference points indexed by t, $R_i(t)$, imply that they need not be static or exogenously given; instead they can endogenously evolve over time with the experiences of the agents. Based on the current period payoffs, the reference point can be revised up or down. This revision can be symmetric, for example, a simple average of the current period payoffs and the reference point. A richer case as shown in Equation (19) indicates that this revision can be asymmetric; in Equation (19), downward revisions are more pronounced than upward revisions.

$$R_i(t+1) = \begin{cases} R_i(t) + \alpha^+ \ (U_i(t) - R_i(t)), & if \ \ U_i(t) \geq R_i(t) \geq 0, \\ R_i(t) - \alpha^- \ (R_i(t) - U_i(t)), & if \ \ R_i(t) > U_i(t) \geq 0. \end{cases} \tag{19}$$

In the above equation, α^- and α^+ are revision parameters and $\alpha^-, \alpha^+ \in [0,1]$. The case with $\alpha^- > \alpha^+$ would indicate that the agents are more sensitive to negative payoff deviations from their reference points.[12]. For the rest of the simulations in this paper, we have utilized the asymmetric revision case.

4 Simulation Design

4.1 A Summary of the Model

4.1.1 Scale Parameters

Table 1 is a summary of the agent-based Scarf model which we describe in Sections 2 and 3. It also provides the values of the control parameters used in the rest of this paper. The Walrasian dynamics of the Scarf economy run in this model has the following fixed scale: a total of 270 agents ($N = 270$), three goods ($M = 3$), and hence 90 agents for each type of agent, $N_1 = N_2 = N_3 = 90$. The initial endowment for each agent is 10 units ($w_i = 10$), which also serves as the Leontief coefficient for the utility function. With this setting, it is easy to figure out that the competitive equilibrium price will be

$$P_1^* = P_2^* = P_3^* = 1. \tag{20}$$

Our simulation routine proceeds along the following sequence for each market day (period): production process (endowment replenishment), demand generation, trading, consumption and payoff realization, learning and expectations updating, propensity and reference point updating, and learning scheme updating. Each market day is composed of 10,000 random matches ($S = 10,000$), to ensure that the number of matches is large enough to exhaust the possibilities of trade. At the beginning of each market day, inventories perish ($\delta = 1$) and the agents will receive fresh endowments (10 units). Each single run of the simulation will last for 2,500 days ($T = 2,500$) and each simulation series

[12] Results do not vary qualitatively for perturbations of these parameters.

Table 1: Tableau of Control Parameters

Parameter	Description	Value/Range
N	Number of agents	270
M	Number of types	3
N_1, N_2, N_3	Numbers of agents per type (4, 5)	90, 90, 90
w_i $(i = 1, 2, 3)$	Endowment (1)	10 units
	Leontief coefficients (2)	10
δ	Discount rate (Perishing rate)	1
S	Number of matches (one market day)	10,000
T	Number of market days	2,500
$P_i^e(0)$ $(i = 1, 2, 3)$	Initial price expectations (3)	$\sim$ Uniform[0.5, 1.5]
φ	Parameter of price adjustment (12)	0.002
θ_1	Parameter of price adjustment (14)	$\sim$ Uniform[0,0.1]
θ_2	Parameter of price adjustment (14)	1
K	Number of arms (Section 3.3.1)	2
λ	Intensity of choice (16)	[0, 1, 2, 3, ..., 10]
$POP_{a_{il}}(0)$	Initial population of a_{il}	1/2
$POP_{a_{sl}}(0)$	Initial population of a_{sl}	1/2
ϕ	Recency effect (17)	0
α^+	Degree of upward revision (19)	0.1
α^-	Degree of downward revision (19)	0.2

The numbers inside the parentheses in the second column refer to the number of the equation in which the respective parameter lies.

(to be discussed in Section 5.1) is repeated 50 times.[13] These scale parameters will be used throughout all simulations in Section 5.

4.1.2 Behavioral Parameters

The second part of the control parameters is related to the behavioral settings of the model, and begins with initial price expectations and price expectation adjustment, followed by the parameters related to the meta-learning models. As noted earlier, we set good 3 as a numéraire good, whose price is fixed as one. Most of these parameters are held constant throughout all simulations, as indicated in Table 1. Some, however, become the interest of the study. Their values are either specified in Table 1, such as intensity of choice (λ), or will be detailed in the sections in which they are involved, which include the prior distribution of price expectations, $P_i^e(0)$ (Simulation 1B, Section 5.2), as well as the initial population of individual-learning agents, $POP_{a_{il}}(0)$ (Simulation 5, Section 5.6).

We could systematically vary the values of different parameters. The focus of this paper is on the intensity of choice (λ) (Table 1). The initial price vector, $\mathbf{P_i^e(0)}$, for each agent is randomly generated where the prices lie within the range [0.5,1.5], i.e., having the Walrasian equilibrium price as the center. Ex-

[13] We have examined the system by simulating the same treatments for much longer periods and we find that the results are robust to this extended horizon.

cept in the cases involving experimental variations of initial conditions, agents' learning schemes, namely, individual learning (innovation) and social learning (imitation) are initially uniformly distributed ($POP_{a_{il}}(0) = POP_{a_{sl}}(0) = 1/2$). The details regarding the implementation of the simulations are provided in Appendix A.

5 Simulation Results

5.1 A Summary of the Simulations

This paper is composed of the following five simulation series, each with a given purpose, and they together address the contribution of meta-learning to the market mechanism. We begin with the qualitative replications of Gintis' results, which is to show that neither individual learning nor social learning alone is sufficient for coordinating the economy toward the Walrasian equilibrium (Section 5.2). We demonstrate these 'market failures' through Simulation 1A (Section 5.2.1) and Simulation 1B (Section 5.2.2). We also show that the observed deviation pattern is different. As for individual learning, we cannot find the domain of attraction to the competitive equilibrium (CE) or to any limit price; instead, the simulation is filled with 'vicious cycles' and price spirals. For social learning, it always converges to a limit price depending on the initial price distribution, while not necessarily to the CE.

If neither individual learning nor social learning alone can ensure price discovery, then would the combined use of them make anything different and, if so, why? This is the main question pursued by the following simulation series. However, to gain a basic insight into the information spread when social learning is replaced with meta-learning, Simulation Series 2 is designed with a different proportion of general-equilibrium agents (Section 5.3). Here, we find that 'rational expectations' (the expectations of general-equilibrium agents) cannot fully spread out to the whole economy and how well it can spread depends on the initial population of the general-equilibrium agents. The endogenously formed blocks which prevent information spread will be further addressed in the following simulation.

The meta-learning model with different intensities of choice is studied in Simulation Series 3; there we find a discernible relation between price convergence and intensity of choice (Section 5.4). We find that the convergence result is also correlated with the endogenously-generated market fraction. Motivated by the observed correlation between price deviation (convergence) and market fraction, our subsequent simulation series attempts to understand the possible causal mechanism between intensity of choice and price deviation. In this regard, Simulation Series 4 treats market fraction as an exogenously given variable and examines whether it alone can be a sufficient causal factor for price convergence, termed as the *market fraction hypothesis* (Section 5.5), whereas the Simulation Series 5 takes market fraction as endogenously formed and studies its path-dependent property (Section 5.6).

Simulation Series 4 is then an auxiliary simulation attempting to see whether the significance of the market fraction can be exogenized, by directly imposing a fixed proportion of innovators, who follow only the individual learning scheme, and imitators, who follow only the social learning scheme (Section 5.5). The general failure to have price convergence to the CE with an exogenous treatment of the market fraction compels us to continuously focus on the endogenously formed market fraction in Simulation Series 5 (Section 5.6). We find the existence of a path-dependence of the market fraction, and this property is also dependent on λ.

To further understand the relationship among the intensity of choice, market fraction and price deviation, we first examine the endogenously formed market fraction from the viewpoint of resource allocation in terms of a balance between exploitation (individual learning) and exploration (social learning). We ask whether the resulting market fraction can be economically justified. To do so, we check payoff equality between imitators (social learning) and innovators (individual learning) (Section 5.7). Interestingly, payoff inequality between these two types of agents exists and is both extensive and persistent; obviously, the market fraction fails to adjust itself to narrow the payoff inequality gap. This provides evidence for the misallocation between exploitation and exploration, which, to some extent, explains the observed price deviation.

To further trace why such an ill-structured market fraction could emerge, we find that the mobility of agents in the sense of their switching between imitators and innovators is key, which, not surprisingly, is determined by the intensity of choice (Section 5.8). The journey along this series of investigations allows us to conclude that the market mechanism can be regarded as a joint force of exploitation and exploration. Each strategy by itself suffers a degree of imperfection, which we can refer to as a ‘type-I error’ and ‘type-II error’. Using payoff data to innovators and imitators, we can see how exploration can aid the performance of exploitation, which may in turn aid the price discovery capability of the entire market economy (Appendix C).

5.2 Simulation 1: Individual Learning and Social Learning Alone

In order to understand the role of learning, we first examine whether an economy composed of innovators (a_{il}) or imitators (a_{sl}) alone can ensure that the agents achieve price coordination.

5.2.1 Simulation 1A: Economy with Individual Learning (Innovators) Only

Simulation 1A studies the effect of individual learning only, which corresponds to Table 1 on the parameter setting that $POP_{a_{sl}}(t) = 100\%, \forall t$. The key result is shown in Figure 1, where we can see that the prices of both commodities 1 and 2 keep on deviating away from the competitive equilibrium prices, which are all one (20). This diverging (flying-away) pattern is quite consistent with respect to different initial conditions, and can be understood as follows.

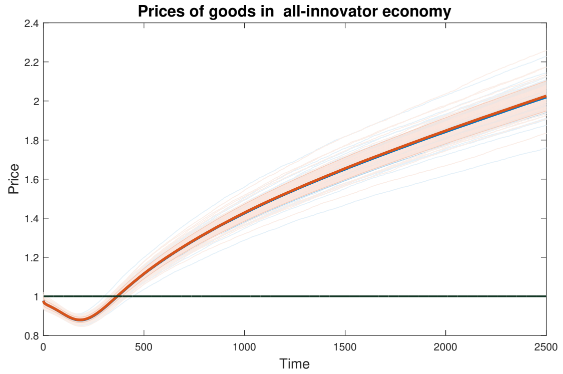

Fig. 1: Price Dynamics with Individual Learning

Agents do not succeed in coordinating to equilibrium prices when all agents are innovators. The prices reported are mean prices over 50 repetitions.

In this economy, based on the gradient-descent algorithm (Equations 9 and 11), agents with unsatisfied demand will upwardly adjust the price expectations of consumption goods and downwardly adjust the price expectations of production goods. However, agents holding the numéraire goods can only upwardly adjust the price expectations of consumption goods, since the price of the numéraire goods is fixed. Either adjustment will scale down the budget constraint and reduce the demand for the consumption goods. This individual contraction, as a group, can cause a mild or serious deficiency in demand, which may further lead to the excessive supply of other agents, even though the matching rate can be improved with adjusted expectations. These agents with excess supply have to downwardly adjust the price expectations of production goods and upwardly adjust the price expectations of consumption goods (Equation 11). This reinforcement can trigger a contraction cycle, which causes the prices in terms of the numéraire to spiral up, as shown in Figure 1.

The numéraire plays a non-trivial role in influencing convergence in our model by introducing a behavioral rigidity given that the price of this commodity cannot be altered[14]. The reason that we can have only explosive price dynamics is because all commodities receive more upward-adjustment potential than downward-adjustment potential (2/3 vs. 1/3 of the market participants), except for commodity 3, which serves as a numéraire. In this manner, there is a net pulling force for the prices of commodities 1 and 2 to spiral up. However, corresponding to Equation (11), there is an reverse process ongoing as well (Equation (13), in general satisfying Equation (13), which is harder than satisfying Equation (11). The set of $\mid \Delta x_{i,j}(t) \mid > 0$ is much *denser* than

[14] Numéraire normalized processes and their convergence properties have been studied widely in the tatônnement literature. We do not analyze the non-normalized case in this paper. See Kitti (2010), for example, on non-normalized iterative processes and the associated convergence conditions.

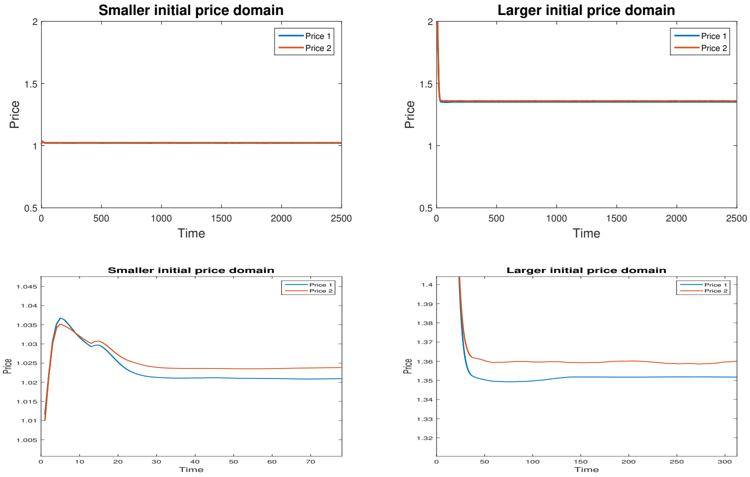

Fig. 2: Price Dynamics with Social Learning

The figures show the simulation results of Simulation 1B-1 (the upper left panel) and 1B-2 (the upper right panel). The initial distribution of priors in Simulation 1B-1 is $U[0.5, 1.5]$, whereas in Simulation 1B-2 it is $U[0, 5]$. The prices reported are mean prices over 50 repetitions. To have a better vision of the transition dynamics, an expanded version of the early periods of 1B-1 and 1B-2 are shown in the lower left and lower right panels, respectively.

the set of $\mid \Delta x_{i,j}(t) \mid = 0$; hence we rarely see the divergence developing in a downward manner.[15]

5.2.2 Simulation 1B: Economy with Social Learning (Imitators) Only

In the next series of simulations, Simulation 1B, we consider the case where all agents are equipped with the social learning scheme *only*, i.e., throughout the entire simulation, they can only learn from others' experiences as described in the protocol given in Section 3.2. This corresponds to Table 1 on the parameter setting that $POP_{a_{sl}}(t) = 100\%, \forall t$. As we shall show in this section, for an economy with only imitators, the results are more subtle. In this case, agents can always coordinate themselves well. However, the limit prices depend on the initial distribution of price expectations $P^e_{i,j}(0), (j = 1, 2, 3)$. Notice that these expectations are initially randomly drawn from a uniform distribution. From Table 1, we know that the range of the initial price distribution is $P^*_j(= 1) \pm 0.5$. With this range, Figure 2, the top left panel at first glance may seem to show that social learning can drive the market to rapidly converge quite close to the Walrasian competitive equilibrium prices.

[15] This property can also been found in Table 2, the results of Simulation Series 3, where we can see that for each type of agent the price expectations of own consumption goods are biased upward, whereas the price expectations of own production goods are biased downward.

However, to see the robustness of this convergence result, we consider a different set of initial price expectations in which the range is not only wider but the centroid is also biased upward from the Walrasian competitive equilibrium price. In other words, we test the price coordination capability when agents' priors are not only more heterogeneous but also more biased. This series of experiments, coded 1B-2, is to be distinguished from the former 1B-1. In Simulation 1B-2, the uniform distribution is set to the interval $[0, 5]$ (with a centroid of $2.5 > 1 = P_j^*$). The result is demonstrated in parallel in the right panel of Figure 2. What we find here is that even though the market can coordinate an 'equilibrium price, the discovered equilibrium was built upon 'wrong' expectations (a Pareto inferior equilibrium). It is also interesting to notice that with this alternative initial distribution the market began with upward biased expectations, i.e., a centroid of 2.5, but it can self-correct this biasedness by bringing it down to a level of 1.3, although the trap to 1.3 shows that this self-correction mechanism is incomplete.

5.2.3 Concluding Remarks of Simulation Series 1

From Simulation Series 1, we can make the following remarks regarding our basic understanding of the role of learning in the agent-based Scarf economy.

Remark 1 Individual learning (innovation) or social learning (imitation) alone cannot ensure price coordination to the Walrasian competitive equilibrium in the agent-based Scarf economy. In an all-innovator economy, the breakdown of coordination from an out of equilibrium configuration is independent of the initial distribution of price expectations. For all-imitator economies, the success of coordination is crucially dependent on the initial distribution.

These findings are in line with the previous literature (Gintis, 2007) that innovation (trial and error learning) and imitation are both necessary to help achieve coordination in a decentralized set-up. Given this, in Section 5.4 we shall examine whether a combination of individual learning (innovation) and social learning (imitation) can lead to price coordination in the presence of a meta-learning model specified in Section 3.3. The meta-learning model introduces agents who can switch in between the individual learning scheme (exploitation) and the social learning scheme (exploration). However, before we examine the impact of meta-learning agents on price coordination, we shall first look at the situation where some agents behave as 'general-equilibrium' agents.

5.3 Simulation 2: Introduction of 'General-Equilibrium' Agents

In economies with only imitators, we saw that the competitive equilibrium prices or those close to them would be eventually adopted by all agents. However, in the meta-learning model, agents could choose not to imitate others if they don't want to. This introduces a problem of the spread of socially desirable

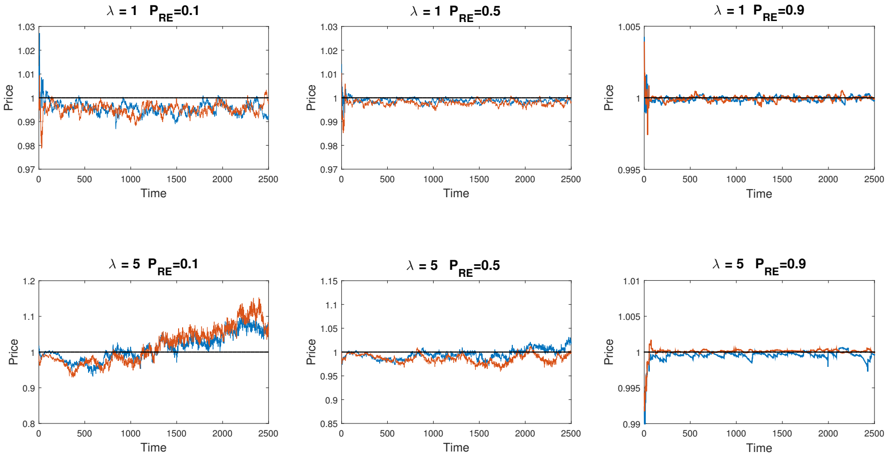

Fig. 3: Price Dynamics with General-Equilibirum Agents

λ indicates the level of intensity of choice and P_{RE} indicates the percentage of general-equilibrium agents. The upper block shows the effect of increasing the general-equilibrium agents from 10% (left panel), 50% (middle panel) to 90% (right panel) for a lower intensity of choice ($\lambda = 1$) and the lower block shows the same effect under a higher intensity of choice ($\lambda = 5$).

expectations. The purpose of Simulation 2 is to evaluate the possible impact of information spread when social learning is replaced with meta-learning. To make the evaluation, we introduce the *general-equilibrium agents*, who take the equilibrium prices as their private prices, remain stubborn and do not alter them in the course of the market dynamics. These agents can be viewed as being the rational-expectations agents, 'fundamentalists' or informed traders. Simulation 2 will examine whether the presence of general-equilibrium agents can drive all agents to adopt competitive equilibrium prices as their private prices.

The key control parameter in Simulation 2 is the percentage of the general-equilibrium agents in the economy, denoted by POP_{RE}, or briefly, P_{RE}, where 'RE' stands for rational expectations. We vary the percentage of general-equilibrium agents from low ($P_{RE} = 10\%$), medium (50%), to high (90%) values. These general-equilibrium agents are distributed equally across all different types of agents. In addition to P_{RE}, we also vary the intensity of choice (λ) from low ($\lambda = 1$) to high ($\lambda = 5$) so that the effect of general-equilibrium agents can be studied in light of different meta-learning behavior. λ itself is an important parameter in agent-based computational economics (Chen, Chang and Du, 2012), and we shall study its role alone more thoroughly in Section 5.4.

The simulation results are shown in Figure 3. We find that an increasing presence of general-equilibrium agents does result in higher levels of coordination. Figure 4 shows that the prices of different goods fluctuate around the competitive equilibrium price ($\mathbf{P}^* = \mathbf{1}$), and that degree of deviations away from the equilibrium decreases from few cents to nil with increasing percentage of general-equilibrium agents. In addition, the tendency to converge appears to be affected by λ. For example, in the case of $P_{RE} = 10\%$, the deviation away from the equilibrium ranges within 5% to 10% when $\lambda = 5$, whereas it decreases to 1% only when $\lambda = 1$.

The presence of general-equilibrium agents does not automatically ensure that all agents adopt equilibrium prices in a meta-learning model. Their influence has a limit, even though increasing proportions of these agents do enhance the overall tendency of the economy to remain closer to the equilibrium prices. We shall return to this issue at the end of this section after a thorough study of the role of the intensity of choice in the meta-learning model.

5.4 Simulation 3: Intensity of Choice and Meta-Learning

Intensity of choice (λ) has been observed to be a key parameter in determining aggregate outcomes in heuristic switching models (Brock & Hommes, 1997; Brock and Hommes, 1998; Hommes, 2006; Chen, Chang and Du, 2012). In Simulation Series 3, we simulate the Scarf economy with λ ranging from 0 to 10 with an increment of 1 (Table 1). The results are presented in both Table 2 and Figure 7. We find that there is a strong tendency to coordinate for smaller values ($\lambda \leq 4$) and the prices stay within a three-percent neighborhood of the equilibrium in the aggregate level (see Table 2). With an increasing λ, there is a breakdown in the tendency to coordinate and the prices diverge away from the equilibrium level. Figure 4 demonstrates the percentage of agents converging to the 10% neighborhood of the Walrasian equilibrium price (WE) of Good 1 for different values of λ.[16] For lower values of intensity of choice ($\lambda = 1, 2$), we find that the majority of agents converge. However, this coordination to the vicinity of the equilibrium breaks down for larger values of λ. Figure 5 presents the dynamics of mean prices of good 1 and the associated one standard deviation value across all agents. This, again, reinforces our earlier observation that the economy converges under meta-learning, however, only for very low values of λ. Figure 6 shows the distribution of the mean price of good 1 across all agents in the economy for $\lambda = 0$ (top panel) and $\lambda = 8$ (bottom panel) at different points in time (t=1,1000 and 2000) during the simulation. In all three figures, the data are pooled from 50 different runs of the simulation for each λ. Thus, we find that despite the presence of meta-learning, the economy converges only under some conditions, i.e., for lower values of intensity of choice, and fails to do so otherwise.

A clearer picture emerges from Figure 7, where the deviations of actual prices from the equilibrium level are related to λ and this relationship is non-

[16] The behavior of the price of good 2 is qualitatively similar.

Table 2: Prices of Goods in the Agent-Based Scarf-Economy: Aggregate and Type-Wise

Prices	Aggregate		Type 1		Type 2		Type 3	
$\lambda\downarrow$	$\bar{P}^e_1$	$\bar{P}^e_2$	$\bar{P}^e_1$	$\bar{P}^e_2$	$\bar{P}^e_1$	$\bar{P}^e_2$	$\bar{P}^e_1$	$\bar{P}^e_2$
0	0.997 (0.001)	0.997 (0.001)	0.979 (0.001)	1.006 (0.001)	1.006 (0.001)	0.979 (0.001)	1.007 (0.001)	1.007 (0.001)
1	0.991 (0.002)	0.991 (0.002)	0.965 (0.002)	1.004 (0.002)	1.004 (0.002)	0.965 (0.002)	1.005 (0.002)	1.005 (0.002)
2	0.979 (0.006)	0.978 (0.006)	0.934 (0.006)	0.999 (0.005)	1.000 (0.005)	0.933 (0.005)	1.004 (0.006)	1.003 (0.005)
3	0.985 (0.013)	0.983 (0.013)	0.903 (0.014)	1.014 (0.013)	1.015 (0.012)	0.900 (0.016)	1.036 (0.016)	1.036 (0.015)
4	1.030 (0.026)	1.030 (0.026)	0.869 (0.028)	1.077 (0.027)	1.077 (0.027)	0.867 (0.028)	1.143 (0.030)	1.146 (0.030)
5	1.152 (0.061)	1.152 (0.061)	0.842 (0.054)	1.250 (0.069)	1.250 (0.070)	0.834 (0.051)	1.366 (0.072)	1.374 (0.074)
6	1.375 (0.113)	1.366 (0.113)	0.866 (0.089)	1.533 (0.151)	1.548 (0.151)	0.881 (0.093)	1.710 (0.125)	1.685 (0.131)
7	1.455 (0.114)	1.462 (0.114)	0.843 (0.077)	1.632 (0.131)	1.629 (0.149)	0.840 (0.078)	1.893 (0.191)	1.914 (0.170)
8	1.444 (0.122)	1.434 (0.122)	0.816 (0.115)	1.560 (0.152)	1.564 (0.146)	0.825 (0.103)	1.950 (0.195)	1.916 (0.187)
9	1.386 (0.111)	1.342 (0.111)	0.772 (0.110)	1.418 (0.133)	1.439 (0.135)	0.782 (0.106)	1.946 (0.215)	1.825 (0.258)
10	1.342 (0.112)	1.291 (0.112)	0.753 (0.107)	1.329 (0.115)	1.355 (0.115)	0.763 (0.101)	1.917 (0.232)	1.781 (0.222)

The above table shows the mean price expectations of the whole economy (columns 2 and 3) and the mean price expectations held by each type of agent (columns 4 to 9), under different λs. To obtain these numbers, the mean price expectations of each run is derived first: $\bar{P}^e_{j,run} = \sum_{t=2,001}^{2,500} \bar{P}^e_{j,run}(t)/500$ ($j = 1, 2$), then averaged over 50 runs, $\bar{P}^e_j = \sum_{run=1}^{50} P^e_{j,run}/50$. The number inside the parentheses shown below each $\bar{P}^e_j$ is the standard deviation of $P^e_{j,run}$, $(run = 1, ..., 50)$.

linear. For smaller values of λ, there is a clear positive relationship between intensity of choice and the *mean absolute percentage error* (MAPE) of the price of good 1. For $\lambda > 5$, the deviation is no longer monotonic and mean prices for different repetitions fluctuate at around 30-45% higher than the equilibrium level.

Although we have identified the role of intensity of choice in driving convergence, we have not precisely identified the channel through which this parameter manifests itself. The following subsections will be devoted to addressing this issue.

5.5 Simulation 4: Market Fractions

The Scarf economy with the meta-learning model allows us to look at the mesoscopic structure or the market fraction by referring to the percentage of agents who chose to be innovators (employing the social learning scheme) and imitators (employing the individual learning scheme) on each market day.

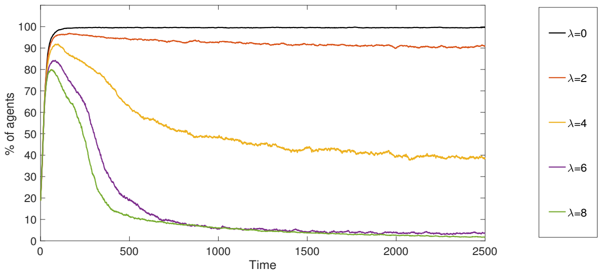

Fig. 4: Percentage of agents converging to Walrasian equilibrium

The above figure indicates the percentage of agents who converge to the 10% neighborhood of the Walrasian equilibrium price (WE) of good 1 for different values of λ. The data are pooled from 50 different runs of the agent-based Scarf economy for each value of λ. For lower values of intensity of choice ($\lambda = 1, 2$), the majority of agents converge. However, this coordination to the vicinity of WE breaks down for larger values of λ.

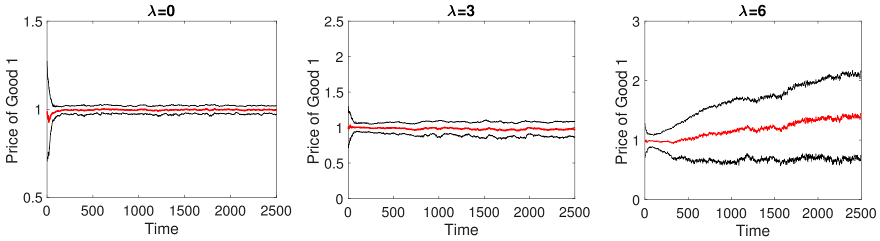

Fig. 5: Price dynamics for good 1

The above figure presents the dynamics of the mean price of good 1 (in red) across all agents in the economy for different values of λ. The dotted black lines denote the one standard deviation from the mean. The data are pooled from 50 different runs for each value of λ. The convergence of mean prices to the Walrasian equilibrium is observed for lower values of λ and it breaks down as λ increases.

Denote this profile as

$$\{POP_{a_{il}}(t), POP_{a_{sl}}(t)\},$$

where $POP_{a_{il}}(t)$ and $POP_{a_{sl}}(t)$ are the number of innovators and imitators on market day t, respectively. The market fraction of innovators $(mks_{a_{il}}(t))$ can then be defined as the proportion of innovators in the total population. In this paper, the term market fraction is reserved for the market fraction of innovators.

The market fraction is not given exogenously but evolves endogenously. By varying the micro-level parameter governing learning (more precisely, λ), we can observe its impact on the meso and macro phenomena. We can then

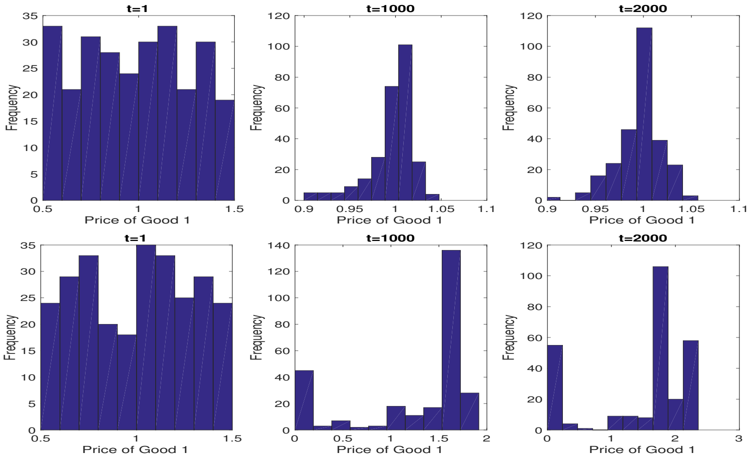

Fig. 6: Distribution of prices overtime

The above figure presents the distribution of the mean price of good 1 across all agents in the economy for $\lambda = 0$ (top panel) and $\lambda = 8$ (bottom panel) at t=1,1000 and 2000. The data are pooled from 50 different runs for each value of λ. The mean prices are seen to converge to the neighborhood of the Walrasian equilibrium for a low value of λ and fail to do so for the higher value.

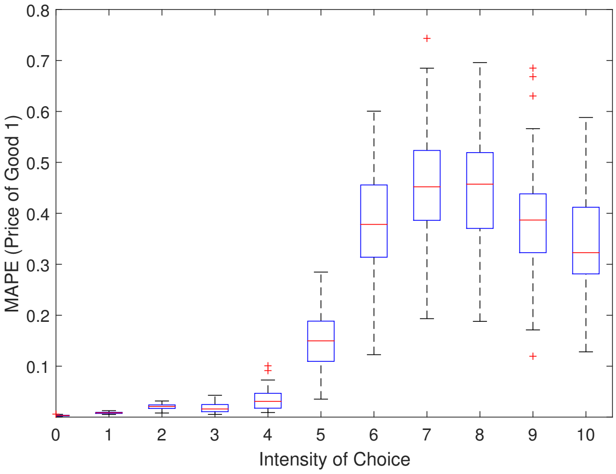

Fig. 7: Price Deviation and Intensity of Choice

The price deviation, away from the Walrasian competitive equilibrium P_j^*, is defined by the mean absolute percentage error (MAPE), which is calculated using the market price of the last 500 market days in each single run, i.e., $MAPE(P_j) = \frac{1}{500}\sum_{t=2,001}^{2,500} \mid P_j(t) - P_j^* \mid$ $(j = 1, 2)$. Since we carry out 50 repetitions for each λ ($\lambda = 1, 2, ..., 10$), the 50 $MAPE$s with respect to each λ are drawn from the minimum to the maximum in a box plot. The figure above only gives the MAPE of good 1 ($j = 1$), which nonlinearly increases with the intensity of choice (λ).

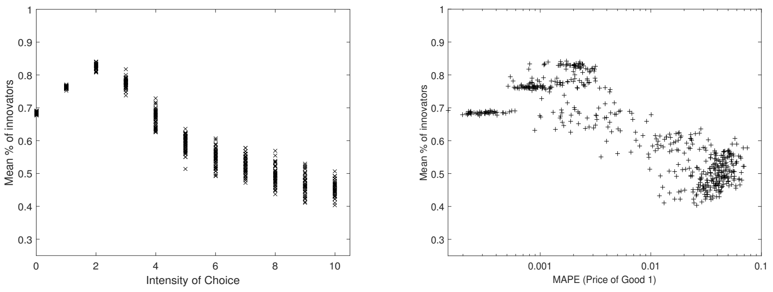

Fig. 8: Market Fractions, Intensity of Choice and Price Deviations

The figure above shows the relationship between market fractions and intensity of choice (upper panel), and the relationship between market fractions and price deviations (lower panel). The data used here is from Simulation 3 (Section 5.4), the same data used to make Figure 7. What are added here are the corresponding market fractions, which, as price deviations in Figure 7, are also derived from the average of the last 500 days, i.e., $\overline{mks_{a_{is}}} = (\sum_{t=2,001}^{2,500} mks_{a_{is}}(t))/500$. Since we carry out 50 repetitions for each λ, the 50 market fractions with respect to each λ is drawn from the minimum to the maximum in the upper panel. On the lower panel, we give the X-Y plot of the pair of the market fraction and the MAPE in the same run by pooling together all these pairs from different λs, a total of 550 (=55 × 10) points. To enhance visibility, MAPE is shown in the logarithmic scale.

speak of the market fractions that are *associated* with converging and diverging prices at the macro level.

Figure 8 shows the second part of the results of Simulation 3. We find that there is a non-linear structure to the relationship between the market fraction (the meso feature) and the level of price coordination (the macro phenomenon). First, the relationship between intensity of choice and the resulting market fractions is summarized in the upper panel of Figure 8. We can see that there is an initial increase in the market fraction of innovators in an economy as the intensity of choice increases. However, for $\lambda > 2$, this trend is reversed and the market fraction of innovators falls and hovers around 50% for different repetitions[17]. Second, concurrently, we also observe that the price deviations are negatively related to the market fraction of innovators in an economy (see Figure 8, the lower panel). Smaller price deviations from equilibrium levels are associated with higher levels of innovators (70-80%). Figure 8, the lower panel, also shows that the relation between price deviations and the percentage of innovators reveals two broad clusters, associated with lower and higher values of λ.

[17] For $\lambda = 0$, past performance should not influence the current choice and therefore $Prob_{i;k}^{t+1} = 1/2$. Thus, we would expect the market fractions to be 50-50, in contrast to what is observed. However, the past performance does exert an indirect influence through the reference point mechanism. Note that the agents consider switching only when the utility falls below their current reference point. Since innovators have relatively high pay-offs for $\lambda = 0$ (and other lower values), the reference point mechanism introduces a bias in favor of the innovators, which explains the deviations we observe in Fig.8.

The inadequacy of learning from experience (innovation) alone to steer the economy into an equilibrium and, consequently, the need for imitation has been pointed out by Gintis (2007). In Simulation Series 1, we have already established that an economy with only innovators or imitators cannot facilitate price coordination to equilibrium levels. We extend this observation by asking whether a combination of innovators or imitators can help explain coordination to the Walrasian equilibrium. Given that high levels of innovators are associated with a tendency for convergent prices under meta-learning, we cannot help but wonder whether market fractions drive price convergence or whether the market fraction alone is *sufficient* to explain price coordination to equilibrium. If this were to be true, then we can make a case for ignoring the micro-level characteristics and focus on the meso structure of the economy alone. This can be formulated more precisely as the 'market fraction hypothesis': an exogenously given market fraction (a linear combination of innovators and imitators) would be sufficient to ensure price coordination in the agent-based Scarf economy.

Note that the market fractions of innovators for $\lambda = 1$ and $\lambda = 3$ are comparable, but their price divergence metrics (MAPE) in 7 are not identical, and instead show an increasing pattern. To test this more rigorously, we simulate the economy where agents have fixed schemes of learning that are preassigned to them. There is no meta-learning and the market fraction of innovators and imitators in an economy is exogenously fixed. This proportion and the set of agents who innovate (or imitate) remains unchanged for the rest of the simulation. This design will enable us to test the validity of the market fraction hypothesis. Simulation Series 4 experiments with different combinations of exogenously given market fractions. Figure 9 reports the result for the case where the fraction is fixed at 70%-30% (the left panel) and 75%-25% (the right panel), the levels corresponding to the instances where convergence was observed in the meta-learning model (see Figure 8, the lower panel). We ran fifty repetitions for each experimental setting and the mean aggregate prices for the final 500 periods are averaged across repetitions. We find that the prices explode (unlike for the meta-learning model) when the market fractions are fixed exogenously (see Figure 9). Even for other alternative fixed combinations (30-70, 50-50, etc.), prices do not converge to the Walrasian equilibrium prices. To understand the underlying mechanism, notice that there are two distinct effects: the stabilizing effect exerted by the imitators and the explosive effect of the pure innovators. The group of imitators among themselves will converge to the initial price held by the agent who is closest to the equilibrium price (simulation 1B). However, the innovators will contribute to the explosive price dynamics for reasons explained in simulation 1A. Under *any* fixed combination, the combined effect will result in non-convergence because the explosive effect of innovators will overpower the stabilizing effect of imitators on the average prices in the economy. This is due to the absence of the information exchange needed for coordination under a fixed linear combination set-up. Hence, we can conclude that meso-level observations alone cannot explain the coordination process.

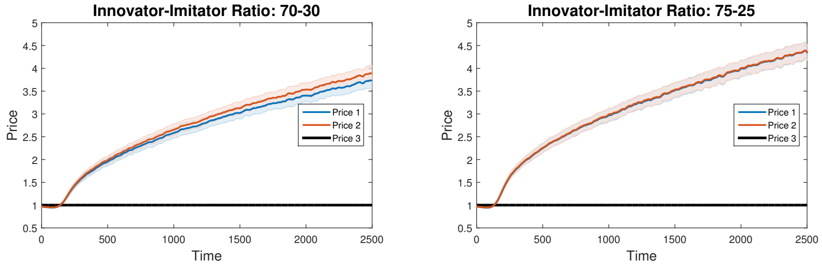

Fig. 9: Price Dynamics of Exogenously-Setting Market Fractions

The above two panels show the price dynamics of 50 repetitions under two fixed market fractions, namely, 70% (the left panel) and 75% (the right panel). In these simulations, 70% (75%) of agents are assumed to be innovators not just initially, but remain innovators throughout the entire duration of the simulation, and similarly for the rest of the agents, who are assumed to be imitators and continue to behave in that way.

5.6 Simulation 5: Path Dependence of the Market Fraction

Section 5.5 shows that the exogenously given (fixed) market fraction alone cannot ensure that the market will coordinate itself at the Walrasian competitive equilibrium; hence, the competitive equilibrium has to be understood as an outcome co-emerging with the market fraction endogenously determined by intensity of choice (λ). However, in addition to λ, coordination can also depend on the initial distribution of the market fractions, $mks_{a_{il}}(0)$. We examine whether the dynamics of the market fractions is *path-dependent* in Simulation Series 5.[18] To provide a better motivation, Figure 10 demonstrates the results of a slice of Simulation 5, i.e., one single run of the agent-based Scarf economy under different initial market fractions ($mks_{a_{il}}(0)$) with respect to the same set of λs used in Simulation 3.

Figure 10 shows the time series plots of the market fractions under different settings. The six boxes, from the upper left to the bottom right, demonstrate the results for different values of λ indicated on top. Within each box, the five series correspond to five different initial values ($mks_{a_{il}}(0)$) as indicated in Figure 10. Regardless of their initial value, the prices tend to evolve toward a narrow corridor. Notice that they all begin with a wider range from 0.1 to 0.9, but end up with a narrower range, a range from 0.1 to 0.25. Hence, in this sense, there is a converging force working to weaken the effect of the initial conditions to some degree.

Nevertheless, depending on the λ, we also see some fundamental differences among the cases with smaller λs ($\lambda \leq 4$) and those with larger λs ($\lambda > 4$). First, the exact location of the formed corridor depends on λ. For the case with smaller λs, the centroids of the corridors seem to have a tendency to move

[18] At this point, $mks_{a_{il}}(0)$ is evenly distributed over two different learning schemes (Table 1).

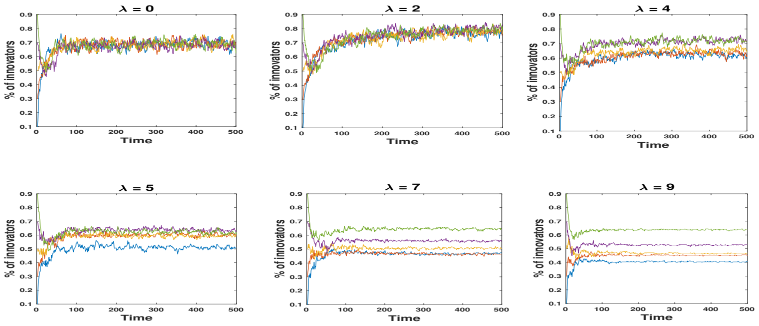

Fig. 10: Time Series of Market Fractions

In the figure above, each box corresponds to a specific λ indicated. Inside each box, there are five time series of market fractions, corresponding to five different initial market fractions ($mks_{a_{il}}(0)$=0.1, 0.3, 0.5, 0.7, and 0.9). Each of the plot of $mks_{a_{il}}(t)$ is based on a single run of the agent-based Scarf economy.

toward a value substantially larger than 0.5, indicating a majority of innovators (agents following the individual learning scheme), whereas for the cases with larger λs, the centroids of the corridors remain at a level generally around 0.5, indicating that innovators and imitators are equally populated (to be discussed further in Table 3). Second, in addition to different levels of centroids, we also notice the differences in the volatility of the market fraction. For the cases with smaller λs, the higher centroid is accompanied by a larger fluctuation of market fractions, implying a high turnover rate in between innovators and imitators. On the other hand, for the cases with larger λs, the market fraction is rather steady, which might suggest a lower turnover rate in between two types of agents. We examine this further in Section 5.8.

To show that the results are not just limited to a single run, in Simulation 5 we conduct 50 runs with the following five market fractions: $mks_{a_{il}}(0)$=0, 0.25, 0.50, 0.75, and 1.00. The results are shown in Table 3 in the Appendix B. In Table 3 we present the mean of the market fractions, and the mean is taken over the last 500 market days. In addition, the volatility of the series is indicated in the parentheses in the form of the standard deviation. From this, we can see that the table can be divided into two parts by λ, namely, $\lambda \leq 4$ vs. $\lambda \geq 4$, and the three salient features revealed in Figure 10 are sustained. First of all, λ can determine who the majority is. When λ is small, innovators globally dominate, but when λ is large innovators dominate only if they dominate in the initial period. Second, from the value inside the parentheses, λ can also determine the degree of variability of the market fraction. The degree of variability is high when λ is small, and it is low when λ is large. Finally, the effect of the initial condition is substantially weakened. It is particularly so for small λs, where the mean effect is independent of the initial market fraction; nevertheless, path dependence does exist for large λs, where the mean fraction increases with the initial fraction.

5.7 Payoff Inequality

Up to this point, we can see that the key parameter in the meta-learning model, λ, has a quite extensive effect on the decentralized market economy, from the price deviations and market fractions. This is not surprising since if decentralization implies a search in terms of exploration and exploitation, then a parameter characterizing the allocation of the resources to these two fundamental economic activities should be critical.

Presumably, we may expect that a good value of λ would allow for a balanced search in a way that the payoffs for various learning schemes are essentially equivalent. If one search scheme (learning scheme), say, exploitation (individual learning), is more efficient, the resources will be reallocated to it from exploration (social learning) and vice versa and, in the end, the payoffs for the two schemes will be equivalent. Hence, we ask: would the market fraction be determined in a way that, regardless of its final level, both learning schemes are rewarded with the same payoffs? Or, alternatively, would the payoffs (economic gains) be equally distributed to innovators and imitators within the society?[19]

To answer this question, we retrieve the data generated from Simulation 3, and examine the impact of λ on the payoff equality or inequality. The results are shown in Figure 11. From the figure, we find little evidence supporting the equality of the payoffs, not even averaged over time, with the case where $\lambda = 4$ being an exception. In other cases, a payoff gap between innovators and imitators is persistent throughout the entire simulation. After a point, the width of the gap remains constant; it neither increases nor decreases. The only feature that distinguishes the nine boxes in Figure 11 is again λ. When λ is small ($\lambda \leq 3$), the payoff to the innovators is larger than the payoff to the imitators. However, there is a slight tendency for the gap to shrink with the increase in λ and then the inequality relation reverses when λ is large ($\lambda > 4$) and the gap widens again with a further increase in λ. The gap becomes particularly wide when λ is as high as 7 or 8, at which point the payoff to the imitators is four times higher that the payoff to the innovators.

By combining these results together with the ones in Section 5.6, we make the following observations: the payoff differential only *weakly* determines who the majority is. When $\lambda \leq 3$, the payoff to the innovators is higher, and the innovators are the majority, whereas when $\lambda \geq 4$, the payoff to the imitators is higher, but the innovators still marginally dominates till λ goes beyond 7.

[19] So far, we have not seen many empirical studies directly devoted to examining the payoff distribution among different heuristics, schemes or strategies in the context of adaptive belief systems or heuristic switching models, neither from the simulation studies, nor from the experimental studies. In this regard, the only study close to us is Bossan, Jann, and Hammerstein (2015), but their adjustment is made at the mesoscopic level (a kind of replicator dynamics), and not at a microscopic level.

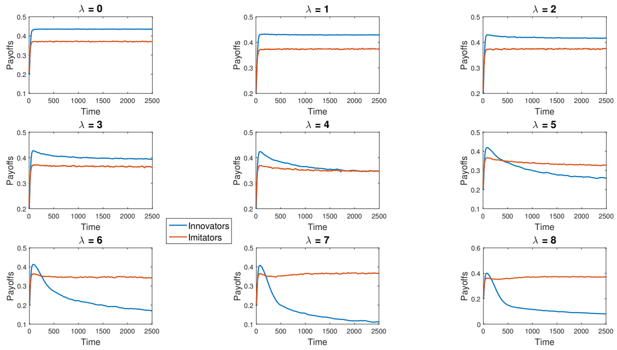

Fig. 11: Payoff Inequality between Innovators and Imitators

The figure is based on the data retrieved from Simulation 3. Each of the boxes above shows the time series of the mean payoffs to innovators and imitators. Each point at time t is calculated as follows. We first figure out the mean of each run: $\bar{U}_{il}(t) = (\sum_{i \in A_{il}(t)} U_i(t))/POP_{a_{il}}(t)$, and $\bar{U}_{sl}(t) = (\sum_{i \in A_{sl}(t)} U_i(t))/POP_{a_{sl}}(t)$, where $A_{il}(t)$ ($A_{sl}(t)$) is the set of agents who applied the individual (social) learning scheme at time t, i.e., the set of innovators (imitators). We then take the mean of these means over 50 runs. The blue line indicates the payoffs to innovators, and the red line indicates the payoffs to imitators. The nine boxes above from the left to the right, and from the upper to the lower levels, correspond to the case λ=0, 1, 2, ...,8.

Except when $\lambda = 4$, there is no instance when λ can deliver equal payoffs.[20] One may wonder why the differences in payoffs can persistently exist. A reasonable conjecture is that, when λ is large, agents have already stopped switching from one scheme to the other scheme due to the *power law of practice* (see Section 5.8 for the details).[21] This is particularly so since in our simulation the recency parameter, ϕ is set to be zero (Table 1), i.e., no forgetting. Hence, some initial good experiences with a given scheme will lock in the agent's choice of it. However, the puzzle remains for the small λs. Do we also expect a similar hysteresis effect there? We shall take a closer look at this issue in Section 5.8.

[20] We are not able to show here the λ which can serve as the equalizer. However, since the payoff reversal happens when λ increases from 3 to 4, we suspect that there may exist some λs ($\lambda \in (3, 5)$) which may remove the gap.

[21] In the psychology literature, the *power law of practice* indicates that subjects' early learning experiences have a dominating effect in their limiting behavior; it is normally characterized by initially steep but then flatter learning curves. In the machine learning literature, it is also known as *premature convergence*, and is a familiar result corresponding to the *path-dependence* property of learning dynamics. In our case, when agents' memory never decays ($\phi = 0$) and λ is large, say, $\lambda = 10$, the path dependency effect can become extreme.

5.8 Mobile Agents

What we have learned up to this stage is that λ not only matters for the determination of price deviations and market fractions (the majority and the minority), but it also matters for the payoffs for different type of agents (the majority and the minority). What is not clear is why the gap in the payoffs among different types of agents can persist simultaneously with rather stable market fractions, specifically for the case where $\lambda \geq 5$ (Figures 10 and 11). Why doesn't the market fraction respond to this gap? On the other hand, it is equally puzzling that even though we see that the market fraction fluctuates over time it has little impact on the gap persistence, specifically for the case where $\lambda \leq 3$ (Figures 10 and 11). To address these questions, we go further down from the mesoscopic level to a microscopic level.

The questions raised above can be partially answered as follows. The provision of two learning schemes or meta learning is only superficial, since some agents never or rarely consider the alternative. In other words, they are rather immobile or nearly immobile. Their prevalence can explain the persistent coexistence of the payoff gap with a stable market fraction. However, we cannot infer the degree of immobility directly from the market fraction statistics. To see this, consider the configuration that half of the agents are innovators and the other half are imitators. In one case, neither of these halves switches to the other scheme, but, in the other case, the same two halves constantly switch at each turn of the market days. In both cases, the market fraction is constantly 50%, but the former implies zero mobility, whereas the latter implies perfect mobility. This simple example shows that Figure 10 can shed little light on agents' mobility. To prove that our conjecture is correct, we therefore need more direct evidence.

The direct evidence can be obtained by tracing the percentage of agents who are immobile in each run. Of course, from a technical viewpoint, unless the probability of switching becomes identically zero (Equation 16), agents will almost surely switch as long as the simulation duration is long enough. Hence, pragmatically speaking, the term 'mobile (immobile) agent' is just a matter of degree, and in this paper we set a minimum threshold, f_{min}, by which agents are termed *mobile agents* if their number of switches during the entire course of simulation, f_i, is greater than this threshold ($f_i > f_{min}$), and are termed *near-immobile agents*, or, simply, *immobile agents* if $f_i \leq f_{min}$.

Figure 12, the left panel, shows the number of immobile agents when f_{min} is set to 1, i.e., an agent will be considered to be an immobile agent if during the entire course of the simulation he has switched at most once. The results reported are the box-and-whisker plot of the 50 repetitions. This figure clearly shows that the number of immobile agents increases with λ. For example, when $\lambda = 10$, from Figure 12 (left panel), 25% to 50% of the agents are nearly immobile. By contrast, when λ is low almost all agents are mobile. Basically, what is revealed here is that even though our meta-learning model allows agents to switch in between innovators or imitators, this flexibility is gone with a large λ. The parameter λ can actually force a large proportion of

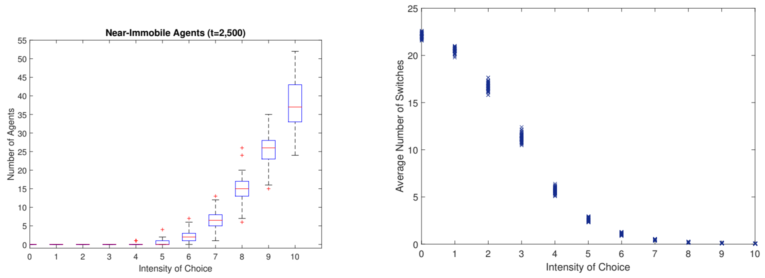

Fig. 12: Near-Immobile Agents and Switching Frequencies

The left panel shows the number of immobile agents with respect to different values of λ. The immobile agent is defined using a switching threshold $f_{min} = 1$. The number shown here is the box-and-whisker plot of the 50 repetitions. The right panel of the figure shows the number of the switches happening on a single market day. A single switch on market day t is observed if an innovator (imitator) at time $t-1$ decides to become an imitator (innovator) at time t. Let $Mo(t)$ be the number of switches observed on market day t. Let $\overline{Mo}$ be the mean taken over the entire course of the simulation, i.e., $\overline{Mo} = (\sum_{t=1}^{2,500} Mo(t))/2,500$. What is drawn in the right panel is the box-whisker plot of these $\overline{Mos}$ from 50 runs under each given λ.

agents to effectively choose to either just be innovators or just be imitators, and then never switch or rarely switch ($f_{min} = 1$).

It is worth mentioning that $f_{min} = 1$ is a very low threshold; the degree of immobility defined accordingly is very conservative. To get a feeling of how immobile these high-λ societies can be, in the right panel of Figure 12 we provide another measure based on the number of switches observed on each market day. As we can see from this panel, the number of switches is almost zero when λ is large. Hence, around 50% of the population only exploits (Table 3), and even though they are moving in the wrong direction, they would not be aware of this or correct their mistakes by learning from others. We can call this a 'Type-I error'. Similarly, around 50% of the population only explores; even though they copy and help distribute wrong beliefs, they never question the validity of these 'public opinions'. We can call this a 'Type-II error'.

The aggregation of these two types of errors reaches its maximum when we have a perfect 'segregation' society, i.e., the high-λ society, of which agents are either innovators or imitators but rarely switch.[22] This explains the earlier observed relationship between price deviation and λ. In fact, a further statistical analysis shows that the correlation coefficient between the number of switches and price deviations (MAPE) is -0.628.[23]

On the other hand, if we look at the low-λ societies, almost all agents will reasonably switch in between the two schemes and hence their exploitations are

[22] As we show in Appendix C, the inferior performance is mainly contributed by innovators rather than imitators.

[23] The correlation is based on the pool of 550 pairs of the MAPE (of good 1) and the average number of switches (over the last 500 periods). There are 550 pairs because we have 50 repetitions for 11 λs.

corrected by explorations and their explorations are checked by exploitations. The results show that when we have roughly 7% (20 out of 270) of the agents who are able to switch per market day, the price deviations are confined to a range of 1% around the equilibrium in the aggregate and within the 5% range for each type (see Table 2).

The microscopic analysis confirms that λ, through its effect on the exploration and exploitation behavior of each agent, can impact the generation and distribution of 'knowledge', in light of what Hayek has emphasized as *the knowledge of the particular circumstances of time and place* (Hayek (1945), p. 521). The market mechanism, in terms of its information aggregation and processing mechanism, can then be substantially different if the ways in which knowledge that is generated and distributed are fundamentally different, caused by the presence of a high percentage of mobile or immobile agents.

In Appendix C, we analyze the relative importance of these two types of errors at the individual level by examining the payoffs to these two types of agents who may contribute to these errors. We find that the market mechanism is a joint function of exploration and exploitation. They help each other, but when one function is not working, exploitation alone can do more harm than exploration alone. This echoes what we have observed in Section 5.2. The reason why exploitation can do more harm than exploration alone is probably because it results in a lower spread of information and hence prevents markets from pooling information effectively.

6 Concluding Remarks

The challenge of Walrasian non-tâtonnement analysis lies in its potentially very large commodity space, which inevitably introduces the dimensionality curse for many modern tools in computational behavioral economics, such as artificial neural nets and evolutionary computation (Chen, Kao and Ragupathy, 2016). Nevertheless, models of learning and adaptation are indispensable in Walrasian non-tâtonnement dynamics. Gintis (2007) had proposed the combined use of individual learning and social learning, both in its simplest form, as an effective recipe. In this paper, we have proposed a meta-learning model to combine both learning schemes into a behaviorally more intuitive framework.

In our meta-learning model, λ (the intensity of choice) turns out to be a key parameter. When λ is low, the Walrasian competitive equilibrium is stable even in the famous Scarf counterexample (Scarf, 1960). The patterns and behavior in economies with a low λ are in stark contrast to those with a high-λ. The difference between these two economies can be understood through their differences in resource allocation between exploration and exploitation as manifested by their market fractions and the agents' mobility. We show that a society composed of mobile agents, who frequently switch between the two learning schemes (innovators and imitators) is crucial for price discovery and the convergence to the competitive equilibrium. We also show that price discovery is achieved through a kind of balance between exploration and exploitation, which concretizes Hayek's 'use of knowledge in the society' (Hayek, 1945).[24] It also focuses on the psychological factors underpinning market outcomes. It is important to note that, in general, there is no price mechanism to coordinate exploration and exploitation.

Although we work with a simple model, many of the features of the present model can be easily generalized (the utility functions, for instance) and additional structures concerning production, technological change and institutions can be introduced. This can help us to understand the dynamics at a decentralized level and can empower us to perform structured computational experiments in complex environments, where the dynamics does not always yield itself to neat, closed-form solutions.

A Implementation details

The details regarding the implementation of the simulations are provided in this appendix. All simulations and analyses were performed using NetLogo 5.2.0 and Matlab R2015a. The NetLogo interface of our simulations is provided in Figure 13. To comply with the current community norm, we have made the computer program available at the OPEN ABM.[25] Figure 13 is the typical NetLogo operation interface. We classify the figure into two blocks. The first block (the left most block) is a list of control bars for users to supply the values of the control parameters to run the simulation. The parameters are detailed in Table 1, including N, M, S, T, φ, θ_1, θ_2, K, λ, POP_{RE} (defined in Section 5.3), and $POP_{a_{il}}(0)$. In addition to these variables, other control bars are the on-off choices for the running model, including individual learning (only), social learning (only), an exogenously given market fraction, and the meta-learning model. For the exogenously given market fraction, $POP_{a_{il}}(0)$ needs to be given additionally.

On the left of the control panels are the real-time demonstrations of the economy under operation. The six subfigures shown in the upper right block are information related to price expectations sustained for a trading day (3) and excess demand and supply settled at the end of a trading day(10). The results are plotted in a time series. The leftmost three subfigures refer to the mean of price expectations (by good and by type), and the middle three subfigures refer to the mean of excess demand and supply (by good and by type).

The top leftmost three subfigures give the summary of the market: prices, quantities, and market fraction (population of innovators). The first one gives the time series plot of the mean of the actual trading prices of goods 1 and 2, denoted by M1_t and M2_t in contrast to its expectations averaged over all agents, M1 and M2 (good 3 serves as the numéraire). The middle one gives the time series of aggregate demand, summed over all agents' planned demand (4). The third one gives the time series of the fraction of agents who adopt the individual learning scheme.

Immediately below the above nine subfigures is the snapshot distribution (dispersion) of price expectations, displayed by goods. The histogram of good 3 is trivial because it serves as the numéraire. The last two subfigures at the bottom provide the information on the relative price of each pair of goods. On the left is the time series of the mean relative price of each pair of goods and on the right is the time series of the respective standard deviation (price dispersion).

[24] See Chen and Venkatachalam (2017) for limits to information aggregation and price discovery in a related context.

[25] https://www.openabm.org/model/4897/

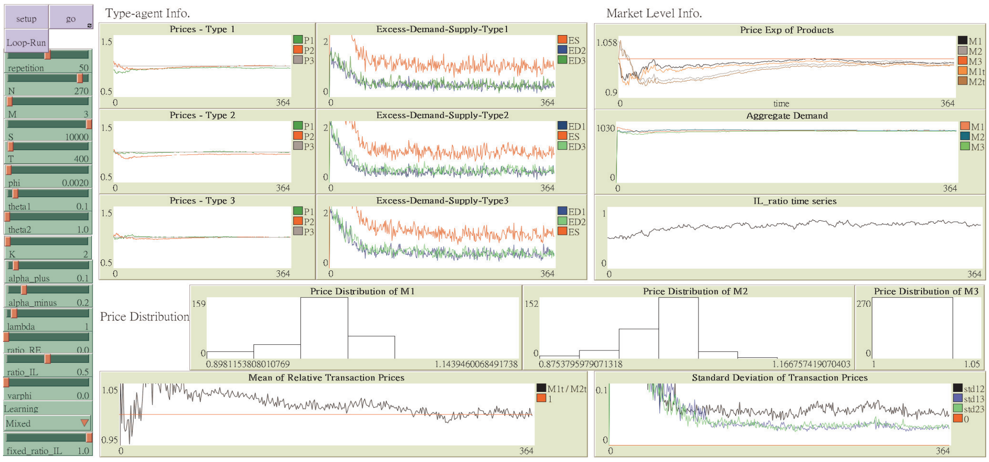
repetition 50
N 270
M 3
S 10000
T 400
phi 0.0020
theta1 0.1
theta2 1.0
K 2
alpha_plus 0.1
alpha_minus 0.2
lambda 1
ratio_RE 0.0
ratio_IL 0.5
varphi 0.0
Learning: Mixed
fixed_ratio_IL 1.0

Fig. 13: NetLogo interface for the simulation of the agent-based Scarf Economy

B Endogenously Determined Market Fractions

This appendix provides the table describing the simulation results concerning endogenously generated market fractions starting from different initial conditions.

Table 3: Endogenously Determined Market Fractions

IIR ↓ / λ	0	1	2	3	4	5	6	7	8	9	10
0	0.684 (0.027)	0.764 (0.025)	0.826 (0.021)	0.764 (0.017)	0.651 (0.011)	0.571 (0.007)	0.515 (0.005)	0.465 (0.004)	0.419 (0.002)	0.377 (0.002)	0.339 (0.001)
0.25	0.685 (0.027)	0.764 (0.025)	0.826 (0.021)	0.769 (0.017)	0.658 (0.011)	0.576 (0.008)	0.523 (0.005)	0.472 (0.004)	0.440 (0.002)	0.420 (0.002)	0.384 (0.001)
0.5	0.685 (0.027)	0.763 (0.025)	0.827 (0.021)	0.775 (0.017)	0.672 (0.011)	0.598 (0.008)	0.551 (0.005)	0.517 (0.004)	0.489 (0.003)	0.470 (0.002)	0.459 (0.001)
0.75	0.685 (0.027)	0.764 (0.024)	0.830 (0.021)	0.781 (0.017)	0.692 (0.011)	0.634 (0.008)	0.594 (0.005)	0.571 (0.004)	0.551 (0.003)	0.546 (0.002)	0.546 (0.001)
1	0.685 (0.027)	0.764 (0.025)	0.832 (0.021)	0.794 (0.016)	0.719 (0.011)	0.677 (0.008)	0.648 (0.005)	0.633 (0.004)	0.630 (0.003)	0.635 (0.002)	0.633 (0.001)

Each cell in the table above corresponds to one parameter value of the initial market fraction (**IIR**), ranging from 0 to 1 with an increment of .25 as shown in the leading column, and one parameter value of λ, ranging from 0 to 10 with an increment of 1 as shown in the leading row. In each pair of the two parameter values, (**IIR**, λ), we simulate the Scarf economy for 50 repetitions, each with a duration of 2,500 days. The mean market fraction of a single repetition is then first derived by taking the average over the last 500 days of each run, i.e., $\overline{mks_{a_{il}}} = (\sum_{t=2,001}^{2,500} mks_{a_{il}}(t))/500$, and then the number reported in each cell of the table is the average of these 50 $\overline{mks_{a_{il}}}$s obtained from the 50 repetitions. As usual, the standard deviation of the series is shown inside the parentheses.

C Payoff Inequality and Two Types of Errors

In Section 5.8, we have seen that large populations of immobile agents associated with large λs cause the market mechanism to malfunction due to both the possible presence of 'type-I' and 'type-II' errors. In this section, we shall have a further look at the relative importance of these two types of errors at the individual level by examining the payoffs to these two types of agents who may contribute to these errors. In Figure 14, we present the results in parallel to Figure 11 except that here we only restrict our attention to those innovators and imitators who are immobile. Since there is only a negligible number of immobile agents when $\lambda < 5$ (Figure 12, the left panel), we only report the payoffs of these two groups for $\lambda \geq 5$ in Figure 14.

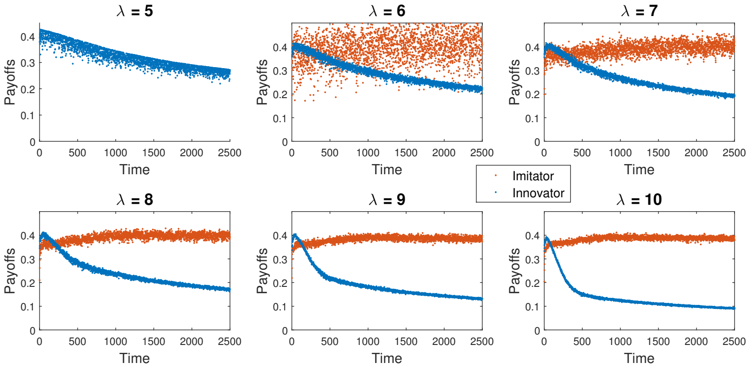

Fig. 14: Payoffs to Immobile Agents: Innovators or Imitators

The figure is based on the data retrieved from Simulation 3. Each of the boxes above shows the time series of the mean payoffs to immobile innovators and immobile imitators. The immobile agent is defined by the given threshold f_{min}. Each point at time t is calculated as follows. We first figure out the mean of each run: $\bar{U}^{f}_{il}(t) = (\sum_{i \in A^{f}_{il}} U_i(t))/\#A^{f}_{il}$, and $\bar{U}^{f}_{sl}(t) = (\sum_{i \in A^{f}_{sl}} U_i(t))/\#A^{f}_{sl}$, where A^{f}_{il} (A^{f}_{sl}) is the set of immobile innovators (imitators). We then take the mean of these means over 50 runs. The blue line indicates the payoffs to immobile innovators, and the red line indicates the payoffs to immobile imitators. The six boxes above, from the left to the right, from the upper to the lower levels, correspond to the case of λ=5, 6,. . . ,10. Notice that the first subplot (λ=5) does not have immobile imitators, and hence the red line is not available. As expected, when the sample size (the population size of the immobile agents) is rather small, we experience a variability in the results, for example, when $\lambda = 6, 7$.

As in Figure 11, Figure 14 shows that, even for the immobile agents, the imitators' performance is superior to that of the innovators. In fact, these two figures together show that the payoffs to innovators drop substantially with an increase in immobile agents, whereas the payoffs to imitators are not affected substantially by the prevalence of immobile agents. Therefore, despite our taxonomy of two types of errors, what contributes to the 'malfunction' of the market mechanism the most is the 'type-one error'. This result demonstrates that the market mechanism is a joint function of exploration and exploitation. They help each other, but when one function is not working, exploitation alone can do more harm than exploration alone.

The above finding resonates well with what we have observed in Section 5.2, in which one economy has only exploitation (Section 5.2.1) and one economy has only exploration

(Section 5.2.2). The reason why exploitation can do more harm than exploration alone is probably because it results in a lower spread of information and hence prevents markets from pooling information effectively. Nevertheless, we have also seen that the performance of innovators (exploitation) can be generally beefed up when learning from others is possible, i.e., being mobile agents. Indeed, when the market is filled with mobile agents (the case with low λs), innovators on average perform better than imitators, as shown in Figure 11 (the sub-figures with small λs).

The above results also shed light on our earlier result that the presence of general-equilibirum agents does not automatically ensure that all agents will adopt equilibrium prices in our meta-learning model (Section 5.3). Why do these 'superior' price expectations fail to spread across the whole economy? The reason is due to the existence of immobile agents who not only block themselves away from the adoption of the 'superior' belief, but may generate lots of 'noises' to prevent others from copying it.

Acknowledgements

We thank the two anonymous referees for their helpful and constructive comments, which have helped us greatly in improving the quality and clarity of the paper. The first and the last author are grateful for the research support in the form of the Ministry of Science and Technology (MOST) grants, MOST 103-2410-H-004-009-MY3 and MOST 104-2811-H-004-003, respectively. We thank Wolfgang Magerl for the able research assistantship in the execution of this project.

References

Albin, P. and Foley, D. (1992). Decentralized, dispersed exchange without an auctioneer: A simulation study. *Journal of Economic Behavior & Organization, 18*(1), 27–51.

Alós-Ferrer, C. & Schlag, K.H. (2009). Imitation and learning. In P. Anand, P. Pattanaik and C. Puppe (Eds.), The handbook of rational and social choice. New York: Oxford University Press.

Anderson, C., Plott, C., Shimomura, K. and Granat, S. (2004). Global instability in experimental general equilibrium: The Scarf example. *Journal of Economic Theory, 115*(2), 209–249.

Anufriev, M. and Hommes, C. (2012) Evolution of market heuristics. *Knowledge Engineering Review 27*(2), 255-271.

Apesteguia, J., Huck, S. and Oechssler, J. (2007). Imitation-Theory and experimental evidence. *Journal of Economic Theory, 136*(1), 217–235.

Arrow, K. (1974). General economic equilibrium: Purpose, analytic techniques, collective choice, *American Economic Review, 64*(3), 253–272.

Arthur, B. (1993) On designing economic agents that behave like human agents. *Journal of Evolutionary Economics 3*(1), 1–22.

Arthur, B. (1993) On designing economic agents that behave like human agents. *Journal of Evolutionary Economics 3*(1), 1-22.

Axelrod R (1997) Advancing the art of simulation in the social sciences, in Conte R, Hegselmann R, Terna P (eds.) *Simulating Social Phenomena*, 21-40. Springer.

Axtell, R. (2005). The complexity of exchange. *The Economic Journal, 115*, F193–F210.

Banerjee, A. V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, 797–817.

Benassy, J. P.(1982). T*he Economics of Market Disequilibrium.* Academic Press.

Bossan, B., Jann, O., and Hammerstein, P. (2015). The evolution of social learning and its economic consequences. *Journal of Economic Behavior & Organization, 112*, 266–288.

Brenner, T. (1998). Can evolutionary algorithms describe learning processes? *Journal of Evolutionary Economics*, 8(3):271-283.

Brock, W. and Hommes, C. (1997). A rational route to randomness. *Econometrica, 65*(5), 1059–1095.

Brock, W. and Hommes, C. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics & Control, 22*(8-9), 1235–1274.

Camerer, C. and Ho, T.-K. (1999). Experience-weighted attraction learning in normal form games. *Econometrica*, 67(4), 827–874.

Chen, S.-H, Chang C.-L, and Du Y.-R (2012) Agent-Based Economic Models and Econometrics. *Knowledge Engineering Review, 27*(2), 187–219.

Chen, S.-H., Kao, Y.-H. and Ragupathy, V. (2016) Computational behavioral economics. In: Frantz, R., Chen, S.-H., Dopfer, K., Heukelom, F., and Mousavi, S. (eds.), *Routledge Handbook of Behavioral Economics*, Routledge, London, 297-319.

Chen, S.-H., and Venkatachalam, R. (2017) Information aggregation and computational intelligence, *Evolutionary and Institutional Economics Review*, 14(1), Quest231–252.

Clower, R. (1975). Reflections on the Keynesian perplex. *Journal of Economics*, 35(1), 1-24.

Ellison, G. and Fudenberg, D. (1993). Rules of thumb for social learning. *Journal of Political Economy, 101*(4), 612–643.

Erev, I. and Rapoport, A. (1998). Coordination, "magic," and reinforcement learning in a market entry game. *Games and Economic Behavior, 23*, 146–175.

Erev, I. and Roth, A. (1998). Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *American Economic Review, 88*(4), 848–881.

Fisher, F. M. (1983). *Disequilibrium foundation of equilibrium economics.* Cambridge, UK: Cambridge University Press.

Gintis, H. (2006). The emergence of a price system from decentralized bilateral exchange. *Contributions in Theoretical Economics, 6*(1), 1–15.

Gintis, H. (2007). The dynamics of general equilibrium. *Economic Journal, 117*(523), 1280–1309.

Gintis, H. (2013), Hayek's contribution to a reconstruction of economic theory, chapter 5 in *Hayek and Behavioral Economics*, edited by Roger Frantz and Robert Leeson, Palgrave Macmillan, pp.111-126.

Grimm, V. and Railsback, S. (2005) Individual-Based Modeling and Ecology. Princeton University Press.

Hahn, F. and Negishi, T. (1962). A theorem on non-tâtonnement stability, *Econometrica, 30*(3), 463–469.

Hayek, F. A. (1945). The use of knowledge in society. *American Economic Review, 35*(4), 519–530.

Hommes, C. and Zeppini, P. (2014). Innovate or imitate? Behavioural technological change. *Journal of Economic Dynamics & Control, 48*, 308–324.

Hommes, C. (2006). Heterogeneous agent models in economics and finance. In L. Tesfatsion and K. L. Judd (Eds.), Handbook of computational economics, Vol. 2 (pp. 1109–1186). The Netherlands: Elsevier.

Hommes, C. (2011) The heterogeneous expectations hypothesis: Some evidence from the Lab. *Journal of Economic Dynamics and Control 35* (1): 1-24.

Hoppitt, W. and Laland, K. N. (2013). *Social learning: an introduction to mechanisms, methods, and models.* New York: Princeton University Press.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Kitti, M. (2010). Convergence of iterative tâtonnement without price normalization. *Journal of Economic Dynamics & Control, 34*(6), 1077–1091.

Koulouriotis, D. E., and Xanthopoulos, A. (2008). Reinforcement learning and evolutionary algorithms for non-stationary multi-armed bandit problems. *Applied Mathematics and Computation*, 196(2), 913-922.

Koza, J. (1992) Genetic Programming: On the Programming of Computers by means of Natural Selection. The MIT Press.

Malinvaud, E. (1977). *The Theory of Unemployment Reconsidered.* Blackwell.

Mandel, A. (2012). Agent-based dynamics and the general equilibrium model. *Complexity Economics, 1*(1), 105–121.

Mandel, A., Landini, S., Gallegati, M., and Gintis, H. (2015). Price dynamics, financial fragility and aggregate volatility. *Journal of Economic Dynamics & Control, 51*, 257–277.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W. et al. (2010). Why copy others? Insights from the social learning strategies tournament. *Science, 328*(5975), 208–213.

Roth, A. and Erev, I. (1995). Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games and Economic Behaviour, 8*, 164–212.

Samuelson, L. (1998). Evolutionary Games and Equilibrium Selection. MIT press.

Scarf, H. (1960). Some examples of global instability of the competitive economy. *International Economic Review, 1*(3), 157–172.

Sutton, R. and Barto, A. (1998). *Reinforcement learning: An introduction*, Cambridge, MA: MIT Press.

Tesfatsion, L. (2006) Agent-based computational economics: A constructive approach to economic theory. In: Tesfatsion L, Judd K (eds.), Handbook of Computational Economics, Volume 2: Agent-Based Computational Economics. North Holland, pp. 831-880.

Tversky, A. and Kahneman, D. (1991). Loss aversion in riskless choice: A reference-dependent model. *Quarterly Journal of Economics*, 1039–1061.

Uzawa, H. (1960). Walras' tâtonnement in the theory of exchange. *The Review of Economic Studies, 27*(3), 182–194.

Velupillai, K. (2015). Iteration, tâtonnement, computation and economic dynamics, *Cambridge Journal of Economics*, 39 (6):1551-1567.

Vriend, N. (2000). An illustration of the essential difference between individual and social learning, and its consequences for computational analyses, *Journal of Economic Dynamics & Control, 24*(1), 1–19.

Wiering, M. and van Otterlo, M. (2012). *Reinforcement learning: State of the art.* Heidelberg: Springer.