The impact of confirmation bias on stock price formation from experimentally-validated cognitive models

Stefano Vrizzi[1,2], Stefano Palminteri[2], Boris Gutkin[1,*], Damien Challet[3,*]

[1] Group for Neural Theory, Département d'Etudes Cognitives, École Normale Supérieure, Paris, France
[2] Human Reinforcement Learning Team, Laboratoire de Neurosciences Cognitives et Computationalles, Département d'Etudes Cognitives, École Normale Supérieure, Paris, France
[3] Laboratoire Mathématiques et Informatique pour la Complexité et les Systèmes, Centrale Supélec, Gif-sur-Yvette, France
[*] Co-last authors

1 Introduction

1.1 Confirmation bias

Extensive evidence shows that individual human behaviour deviates from rationality [7, 8, 39, 40, 80, 81, 83, 84], also specifically in traders and investors [2, 21, 37, 45, 62, 75, 85]. However, mainstream economic theory generally assumes economic actors to be rational [29, 44, 57, 58, 68, 71, 78]. Modelling-wise, the rationality assumption at the cognitive level is convenient, but what are its macroscopic consequences on price? In other words, considering that financial markets aggregate the biased decisions of traders, does the impact of these biased decisions accumulate or cancel out? In this paper, we address this *bias aggregation problem* [5].

Among all human biases, we focus on confirmation bias, which in plain words is the tendency to see only what you already want to see. More formally, it is often defined as a systematic error in belief updating with respect to Bayesian learning [19], but its concept spans over a variety of definitions and nuances [42]. This bias has been extensively documented by psychologists for its pervasiveness in human cognition and social dynamics [18, 25, 28, 35, 50, 61]. Strong evidence of confirmation bias has been gathered in both information acquisition and information use [38]. It is robust to experience and resilient to economics incentives; while incentives manage to encourage Bayesian belief formation, they cannot quench confirmation bias [12]. Further studies, however, find confirmation bias to be less predominant than claimed [26].

Confirmation bias has also been detected specifically in finance: investors preferentially read information supporting their investment decision rather than opposing information [13]. More broadly, commercial investment platforms claim that sticking to the same strategy despite loss is one of the strongest biases in trading. Nevertheless, in finance, confirmation bias has been understudied in comparison to other key cognitive biases, such as the anchoring effect and overconfidence [16].

In our study, we examine how confirmation bias influences market behaviour and individual strategy preferences. Our approach is computational. We simulate stock markets by multi-agent reinforcement learning, where agents learn to forecast the market, trade and price their orders. While the literature already includes studies on multi-agent reinforcement learning models [53–55] and on the impact of confirmation bias [69] in financial markets, we contribute by adding a key missing element: the employment of cognitive models grounded in experimental evidence. The main novelty of the current study is the model adopted for agents' cognition. We leverage experimentally-validated computational models that capture confirmation bias in behavioural experiments, starting from explicit learning mechanisms based on neural evidence. In the next two sections, we introduce and motivate our modelling choices.

1.2 Reinforcement Learning

If we depart from the ideal rationality assumption to study human cognition, we need to opt for a cognitive theory. The debate on the most appropriate one is complex and unsolved [74]. However, financial markets represent an arena where traders' behaviour is clearly goal-driven. They pursue profit, and more generally they aim to learn how the market behaves [23, 41, 43, 51, 60]. Out of the vast scientific literature available, we focus on one prevalent theory: reinforcement learning (RL). In plain words, RL consist in learning which action to take, in a given context, by trial-and-error, when pursuing a goal. An RL agent aims to *reinforce* actions leading to rewards, and discard actions leading to punishments. In other words, the sensitivity to reward and punishment guides the development of the agent's preferences and the adaptation of its behavioural strategy.

RL is grounded in influential empirical evidence from psychology and neurophysiology. Its learning processes have been documented in animal and human behaviour [77, 82], as well as in the dopaminergic neurons in basal ganglia within the human brain [73]. With respect to finance, investors increase their 401(k) savings rate (i.e. a retirement savings plan in the USA) if they personally experience high average and/or low variance return, in line with a reinforcement learning explanation [15]. A further example is that investors increase their participation to IPO auctions when experiencing high returns (however, this may not be the case for institutional investors) [14]. Evidence also encompasses laboratory experiments. In restless bandit-tasks,[1] subjects are Bayesian learners only if nudged into paying attention about payoff shifts, otherwise their behaviour is better explained by RL [67]. With respect to the neurophysiological evidence, the activity of key brain areas associated to reinforcement learning (e.g. ventral striatum) has been linked to trading behaviour [31].

Modelling-wise, both behavioural and neurophysiological aspects of RL can be captured by models building upon a simple delta rule, the Rescorla–Wagner model [70]. In our study, we focus on Q-learning [20, 79].

In short, we choose RL because of its features and its suitability for our purposes. It is computationally explicit, guided by reward sensitivity, goal-oriented, offering a mechanistic explanation of learning and decision-making, and it accommodates traders' adaptation to market conditions.

1.3 Confirmation bias in Reinforcement Learning

RL is also able to capture fundamental distortions in information processing [3, 49, 64, 65]. Among the models available, in this study, we leverage the RELATIVE ASYMMETRIC RL model. We chose this model because it won model comparison in explaining human learning and decision-making in RL tasks from behavioural economics experiments [32, 47]. As the name suggests, this model is based on two modules: RELATIVE and ASYMMETRIC. Here, we focus on the latter, which is based on the gap between expected and observed outcome of an action, i.e. the *surprise* experienced by agent (Fig. 1b). From the RL perspective, rationality is defined as treating equally positive and negative surprises. On the contrary, confirmation bias can be thought of as welcoming positive surprises, while neglecting negative surprises, from chosen actions (and vice versa for unchosen actions). This asymmetry is embodied by the ASYMMETRIC module. Its formal definition of confirmation bias will be provided in Methods.

1.4 Multi-Agent Reinforcement Learning

We address the bias aggregation problem computationally. We simulate a stock market that is fully driven by the decisions of a large number of heterogeneous traders who can learn and adapt their own strategies. Inspired by previous work [11, 30, 36, 46, 53, 56], we develop a stock market simulator in the form of a multi-agent reinforcement learning (MARL) model.

[1] Behavioural economics tasks where contingencies (i.e. underlying rewards associated to one option) change.

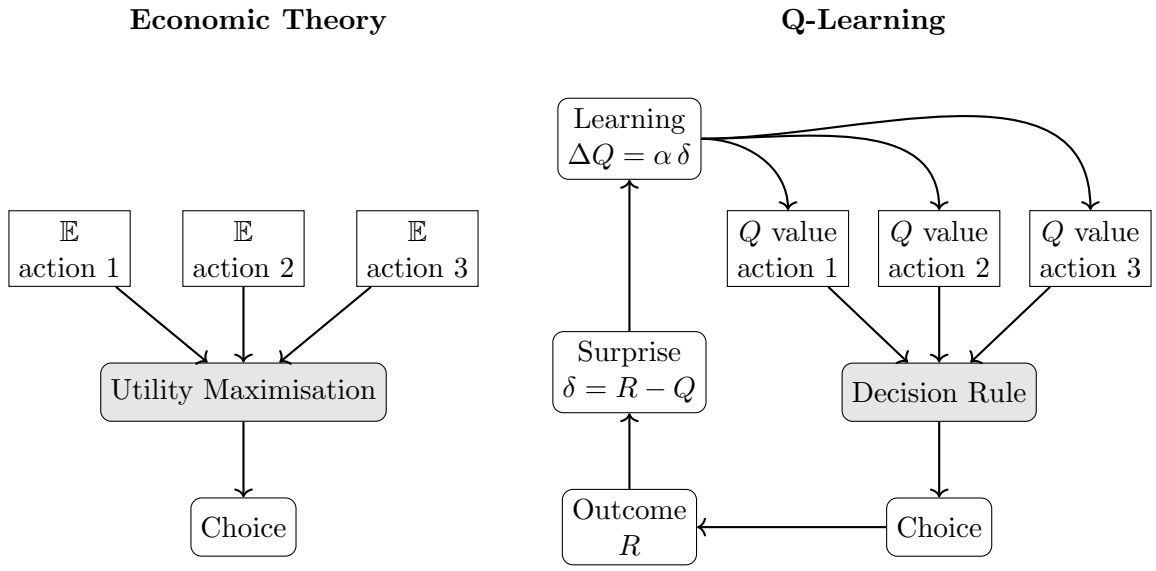

Figure 1: Comparison of cognition models between mainstream economics and Q-learning in RL.
In the economic theory panel, $\mathbb{E}$ denotes subjective expected utility, from subjective utilities weighted by known probabilities of all possible outcomes of that given action. Economic theory may also employ Bayesian learning as a form of inductivism linking subjective probabilities and rationality [17, 24, 34]. In the right panel, we sketch Q-learning as model-free RL, i.e. learning by trial-and-error. The decision rule can be, for instance, argmax, ϵ-greedy, softmax. With respect to learning, for the sake of simplicity, the diagram depicts only the update for the chosen action, but unchosen actions can also be updated, if their outcome is observed (or estimated). Figure adapted from [48].

Since the main model of reference is a MARL called SYMBA (SYstème Multiagent Boursier Artificiel) [52–55], for practical purposes, we refer to our MARL as *pSY*MBA, to reflect the *psy*chological foundations of the model. Among numerous differences, the main novelty with respect to SYMBA lies in the design of agents' cognition:

1. as explained in the previous section, pSYMBA agents employ *human* RL models that have been developed and validated on behavioural economics experiments to capture human biases in learning and decision-making [32, 47, 49, 66];

2. given the paramount importance of expectations in economics [23, 41, 43, 51, 60], decision-making in pSYMBA agents is structured hierarchically upon the forecasting strategy, the upstream decision to form market behaviour predictions that guide and feed into all the downstream decisions (trading and pricing).

2 Methods

We structure our methods section top-down, in two parts: first, the architecture of the financial market simulator (section 2.1), then the cognitive model of individual agents (section 2.2).

2.1 Financial market simulator architecture

2.1.1 Simulation-wide parameters

We simulate $I = 10^4$ agents trading a finite number of shares of a single asset. Prices in the market emerge from the aggregate actions of the agents. Simulations last $T = 6500$ time steps,

however, for our main analysis, we discard the first $T_{\text{train}} = 4000$ time steps, where agents are still learning how to trade. We extract the price return statistics of interest from the equivalent of 10 years worth of trading (2500 time steps), assuming $T_y = 252$ business days in a year.

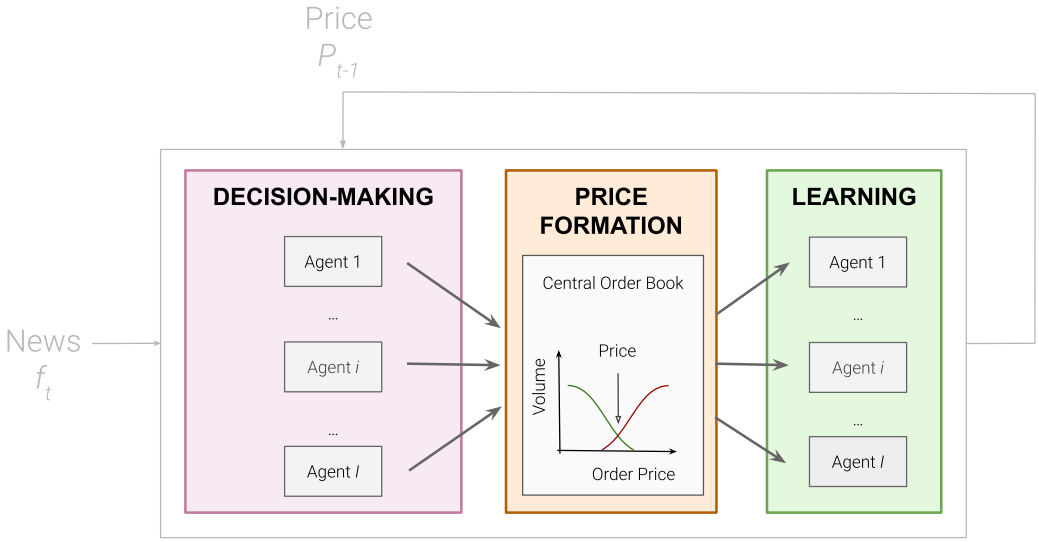

Figure 2: Diagram of information processing during a timestep in pSYMBA. The market is represented as a function producing the stock market price P_t by processing two sources information: exogenous information f_t (the news) and endogenous information (the last price P_{t-1}) from the feedback loop. Stock price results from the interaction of agents' decisions. Decision-making involves forecasting, trading and pricing, for each agent, independently. Their individual trading orders are collected in a central order book, which sets the price through a Walrasian auction. Finally, agents learn from the price outcome.

2.1.2 Exogenous information

We inject exogenous information f into the model to simulate the news from the outside world, which represent the underlying true fundamental value of the asset [4, 56, 63]. We generate a multiplicative stochastic process by geometric Brownian motion (GBM) in the discrete form:

$$\log f_t - \log f_{t-1} = \phi_t, \tag{1}$$

where ϕ_t is characterised by an annual mean $\mu_Y = r_{\text{free}} + r_{\text{premium}}$ (i.e. the desired risk-free and risk-premium assets' annual growth rates), and volatility $\sigma_Y = 0.1$ (Tab. 1).

2.1.3 Price formation

Individual trading orders are sent to a central order book. pSYMBA assigns a timestamp to each order, adding a random delay sampled uniformly $\mathcal{U}(0, 1)$ within the current time step. The central order book sorts bid orders in descending fashion, and ask orders ascending fashion, according to their price, while giving priority to the earlier timestamps in case of equal price. After collecting the trading orders on the central order book, stock price is formed through a Walrasian auction [22]. The auction follows the rules from [27], as described in [72], to determine the price P_t. Once the price is set, the central order book clears trading orders, also considering timestamp priority. Orders may be filled only partially. Trading orders are cleared with a broker fee $b = 0.1\%$.

Name	Value	Purpose
I	10000	Number of agents
S	200	Number of testing simulations per learning condition
T_{test}	2500	Time steps for analysis
r_{free}	1%	Annual risk-free rate on available margin (interests are awarded daily)
r_{premium}	4%	Annual premium rate, i.e. growth rate in addition to r_{free}, characterising the typical (mean) growth rate of the exogenous information f, the true fundamental value
σ_Y	0.1	Annual volatility of exogenous information f (true fundamental value) injected into the model
b	0.1%	Broker fees
tick	0.01	The minimum (upward or downward) amount the stock price can move; currency is arbitrary

Table 1: Summary table of model architecture parameters.
We partition the table in three sections to distinguish: overall input parameters, exogenous information parameters, and market-related parameters.

2.2 Agents' cognition model

2.2.1 Reinforcement learning and confirmation bias

We model agents' cognition by leveraging the RELATIVE ASYMMETRIC RL model. We choose it for empirical reasons: its superior ability in model comparison to explain human behaviour in behavioural economics RL tasks [32, 47, 49, 66].

Agents operate through three Q-learning algorithms based on the RELATIVE ASYMMETRIC model: the forecasting algorithm $\mathcal{F}$ to forecast the stock price, the trading algorithm $\mathcal{T}$ to trade, and the pricing algorithm $\mathcal{P}$ to price trading orders. In the forecasting algorithm, at each time step, each agent can choose either fundamentalism (section 2.2.3) or trend-following (section 2.2.4). In trading, the action menu is: ask, hold or bid; if the agent own no shares, its possible actions are open or wait. In pricing, the action menu includes four types of limit orders, plus market order.

In each algorithm, agents learn the value of each action a in state s according to a simple delta rule $\Delta Q(s,a) = \alpha\,\delta$, where α is the learning rate, while $\delta = R_t - Q_t(s,a)$ is the prediction error between the observed outcome R_t and the expected action outcome $Q_t(s,a)$. R_t is the reward or punishment obtained from a given action, defined by algorithm-specific reward functions (not shown).

In the forecasting and in the pricing algorithm, agents adopt a different learning rate depending on the prediction error sign and on whether the action was or not the one chosen by the agent. Formally, for the chosen action c, the Q update is

$$\Delta Q(s,a_c) = \begin{cases} \alpha_+\,\delta_c, & \text{if } \delta_c > 0 \\ \alpha_-\,\delta_c, & \text{if } \delta_c < 0. \end{cases} \tag{2}$$

The agent employs α_+ when the taken action results in a better-than-expected outcome, while α_- is used when the taken action results in worse-than-expected outcomes. In a mirrored fashion, for an unchosen action u, the Q update is

$$\Delta Q(s,a_u) = \begin{cases} \alpha_+\,\delta_u, & \text{if } \delta_u < 0 \\ \alpha_-\,\delta_u, & \text{if } \delta_u > 0. \end{cases} \tag{3}$$

The agent employs α_- when the unchosen action results in a better-than-expected outcome, while α_+ is used when the unchosen action results in worse-than-expected outcomes. The

learning rate asymmetry defines the learning bias: $\alpha_+ > \alpha_-$ implements *confirmation* bias, while $\alpha_+ < \alpha_-$ corresponds to the opposite condition, *disconfirmatory* bias, which emphasises mistakes, i.e. disconfirmatory information. When the two learning rates are equal, we are back to the basic model with a single learning rate, where all information is treated equally.

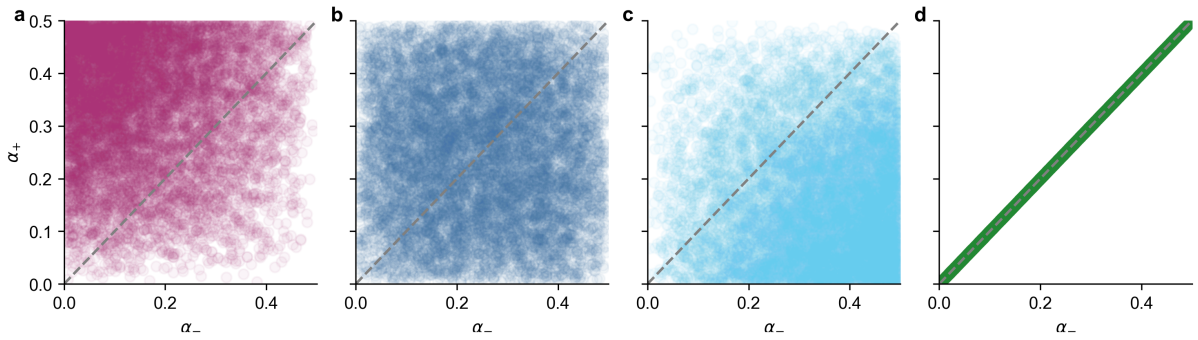

Figure 3: Distributions of confirmatory and disconfirmatory learning rates α_+ and α_- in the agent population, depending on learning condition.
a) Confirmation bias (B_+); **b**) balanced bias (B_0); **c**) disconfirmatory bias (B_-); **d**) rationality (B_{00}).

The trading algorithm is based on expected profits (rather than realised profits), to decide the most appropriate trading action *given the current forecast*. This algorithm does not include any bias, as we are not interested in biasing the coupling between price forecast and trading decision.

2.2.2 Cognitive traits

In addition to the RL learning rates, each agent's cognitive traits are shaped by six agent-specific parameters. Here, we focus on two: investment time horizon τ, or patience, and true fundamental value co-integration speed κ, or simply news sensitivity (Tab. 2). Each of these parameters is sampled independently, for each agent.

Name	Symbol	Purpose	Range
Patience	τ	Number of time steps in the future or back in the past the agent looks at to perform key computations, especially for forecasting and profit expectations; τ also determines the average market-interaction frequency of a trader.	$[T_w, T_y]$
News sensitivity	κ	Speed at which the agent updates its estimate about the exogenous information on the asset; implemented as co-integration speed of an assumed true fundamental value signal (section 2.1.2) that can only be known partially by the agents	$[5 \cdot 10^{-4}, 0.01]$

Table 2: Two key agent-specific cognitive traits: their symbol, purpose, and value range.
Both parameters are sampled uniformly. $T_w = 5$ and $T_y = 252$ indicate, respectively, the number of business days in a week and in a year.

2.2.3 Fundamentalism

Agents have only partial knowledge of the true fundamental value of the asset. They obtain their own estimate by co-integration [59], according to their news sensitivity parameter κ_i:

$$\widetilde{f}_{i,t} = \widetilde{f}_{i,t-1} + \kappa_i(f_t - \widetilde{f}_{i,t-1}). \tag{4}$$

In other words, κ is a measure of the agent's speed in catching the true fundamental value f_t, which changes constantly. We set $\kappa_{\min} = 0.0005$ to allow for minimal opinion dynamics.

2.2.4 Trend-following

We represent chartism by a trend-following estimate from the latest price dynamics. For its chartist forecast, the agent adds a linear projection to the last available price P_{t-1}:

$$H_{i,t} = P_{t-1} + m_{\tau_i}\tau_i \tag{5}$$

More specifically, the agent i considers the last τ_i past prices $\{P_{t-\tau_i}, ..., P_{t-1}\}$, employing the time indexes $k = \{-\tau_i + 1, ..., 0\}$, it centers the last available price P_{t-1} at the origin and fits a linear model $\Delta\widehat{P}(k) = m_\tau\, k$, forcing the intercept to 0.

2.3 Statistical analysis

In the following sections, we address the central question of this thesis: whether the market amplifies or mitigates the impact of individual biases. We first explain our statistical analysis, to then run three main analyses.

2.3.1 Statistical tests

We now conduct statistical tests to assess the differences between conditions: confirmation bias B_+, balanced bias B_0, disconfirmatory bias B_-, rationality B_{00}, and zero-intelligence Z. We label the first four as learning conditions, the latter as no-learning condition.

We run simulations by setting specific a random seed for each instance of pSYMBA. Random seeds differ between training phases and between any testing phase, but they are set to be identical across conditions (including no-learning, i.e. zero-intelligence agents). In other words, exogenous information signals are identical across conditions, and agents are twins, they are characterised by the very same parameters except for the learning rates, which are the independent variables that we want to manipulate. This framework allows us to compare simulations as repeated measures.

We first run a Friedman test, where the null hypothesis is defined as no significant difference between conditions (reported in Tab. 3). We employ a Friedman test rather than a one-way ANOVA because data distributions are not normally distributed. In all Friedman tests, $N = 200$, which is the number of simulations.

If we can reject the null hypothesis ($p < 0.05$), we run a series of Wilcoxon signed-rank tests as post-hoc tests. Again, they replace paired t-tests because data points are not normally distributed. In our Wilcoxon tests, we compare the reference condition to one other condition, as listed in Tab. 3. Each of these tests is a location test, it evaluates if the distribution of the differences between two conditions is symmetric with respect to zero; it does not evaluate if the two distributions differ in shape. We correct p-values by Bonferroni correction, multiplying the p-values obtained from the Wilcoxon tests by the number of pairwise comparisons for the variable of interest (i.e. the number of conditions that we contrast to the reference condition). Since this type of correction is conservative, it increases the probability of false negatives in the post-hoc phase. We will be more confident about detected differences, but less confident about a lack of statistical difference at this stage.

We set all significance thresholds to 0.05.

Analysis	Test	Comparisons	Null Hypothesis $\mathbf{H_0}$
Model validation	Test non-Gaussianity of price returns w.r.t. to Gaussian news	f vs. Z	$M(Z-f)=0$
		f vs. B_+	$M(B_+-f)=0$
		f vs. B_0	$M(B_0-f)=0$
		f vs. B_{00}	$M(B_{00}-f)=0$
		f vs. B_-	$M(B_--f)=0$
Bias aggregation problem	Test changes in price return properties between learning conditions	B_{00} vs. B_+	$M(B_+-B_{00})=0$
		B_{00} vs. B_0	$M(B_0-B_{00})=0$
		B_{00} vs. B_-	$M(B_--B_{00})=0$

Table 3: Summary of statistical analyses.
We apply the above framework to test for differences between conditions in each dependent variable of interest (volatility, skewness, kurtosis). In the last column, M stands for median of the differences between conditions: confirmation bias B_+, balanced bias B_0, disconfirmatory bias B_-, rationality B_{00}, zero-intelligence Z, and the exogenous information f (i.e. the true fundamental value).

2.3.2 Average forecasting strategy choice

We study price dynamics in relation to the dynamics of average forecasting strategy choice $\psi_{\mathcal{F},t}$, averaged across traders (or simply average strategy choice), i.e. the fraction of the total number of agents I choosing fundamentalism ($a_{\mathcal{F}}=1$) at a given time step t:

$$\psi_{\mathcal{F},t}=\frac{1}{I}\sum_i a_{\mathcal{F},i,t}. \tag{6}$$

We derive $\psi_{\mathcal{F},t}$ as the mean RL action in the forecasting algorithm $\mathcal{F}$, averaged across agents I, for each time step t.

2.3.3 Forecasting strategy preference

We define an agent's forecasting strategy preference as the fraction of times that an agent i chooses fundamentalism ($a_{\mathcal{F}}=1$):

$$\phi_{\mathcal{F},i}=\frac{1}{T_{\text{test}}}\sum_t a_{\mathcal{F},i,t}. \tag{7}$$

Values leaning towards zero indicate a preference for trend-following, while values leaning towards one indicate a preference for fundamentalism. We compute this value for each agent, for each testing phase, per learning condition. We then interpolate these values linearly to produce a heatmap of agents' forecasting preferences as a function of τ and κ, for each learning condition (Fig. 5).

3 Emergent properties

Before addressing the central question of this study about the impact of confirmation bias on stock price formation, we analyse the model. The model produces the desired emergent properties across levels. At macroscopic level, it produces stylised facts of price returns; at mesoscopic level, it links agents' forecasting strategy choices to alternating regimes of market efficiency and speculative bubbles; at the microscopic level, agents become either fundamentalists or chartists, in line with their own patience and news sensitivity. We present these findings in the following sections.

3.1 Macroscopic analysis: empirical validation

Trading fundamentally alters the properties of the injected information f (section 2.1.2), giving rise to stylised facts. Price returns from simulations exhibit two key non-Gaussian properties that typically characterise financial markets: negative skewness and excess kurtosis (Tab. 4-5).

Moreover, the (log) variance of price returns also increases, with respect to news (the changes in exogenous information). Although a larger variance is expected when summing two stochastic processes (i.e. the true fundamental value f and agents' choice sampling from softmax decision rule), the size of the additional volatility clearly denotes *excess* volatility [76]. Price return volatility is 14 to 35 times greater than the volatility of the true fundamental value for learning conditions, and 4.4 times for the zero-intelligence condition, in line with foundational empirical literature [76] reporting an excess volatility factor of 5 to 13 times.

Variable	$\chi^2(5)$	p
$\log \sigma^2$	758.768571	9.60×10^{-162}
$Skew$	171.265714	3.92×10^{-35}
$Kurt$	462.714286	8.88×10^{-98}

Table 4: Friedman test comparing exogenous information and price returns from zero-intelligence condition and learning conditions.
A priori comparison to address question 1 in Tab. 3. All Friedman tests are statistically significant, allowing us to run post-hoc comparisons between individual conditions (Tab. 5). P-values are not corrected.

	B_+	B_0	B_-	B_{00}	Z
$\log \sigma^2$	2.65 $p = 7.18 \times 10^{-34}$ $W = 0$	3.23 $p = 7.18 \times 10^{-34}$ $W = 0$	3.58 $p = 7.18 \times 10^{-34}$ $W = 0$	3.32 $p = 7.18 \times 10^{-34}$ $W = 0$	1.48 $p = 7.18 \times 10^{-34}$ $W = 0$
$Skew$	-1.45 $p = 2.99 \times 10^{-22}$ $W = 1966$	-1.11 $p = 1.43 \times 10^{-13}$ $W = 3818$	-0.47 $p = 1.97 \times 10^{-7}$ $W = 5548$	-0.76 $p = 1.76 \times 10^{-8}$ $W = 5210$	-1.19 $p = 9.65 \times 10^{-15}$ $W = 3538$
$Kurt$	33.36 $p = 7.18 \times 10^{-34}$ $W = 0$	30.13 $p = 7.18 \times 10^{-34}$ $W = 0$	15.64 $p = 7.18 \times 10^{-34}$ $W = 0$	24.43 $p = 7.18 \times 10^{-34}$ $W = 0$	32.63 $p = 7.18 \times 10^{-34}$ $W = 0$

Table 5: Impact of trading on news: price returns display stylised facts (excess volatility, negative skewness, and excess kurtosis).
Post-hoc tests contrast each model condition (learning conditions and zero-intelligence) against the exogenous information (Tab. 3). Columns represent confirmation bias (B_+), balanced bias (B_0), disconfirmatory bias (B_-), rationality (B_{00}), and zero-intelligence (Z). Rows represent dependent variables of interest: volatility, skewness and kurtsis of price returns. The reference condition is the exogenous information (f), a geometric random walk (section 2.1.2). Each box shows the results from a pairwise Wilcoxon signed-rank test between the column condition and the exogenous information. The top value is the median difference between the two conditions, subtracting the reference condition f from the column condition. p-values are Bonferroni-corrected by the number of pairwise comparisons (here 5). The bottom value reports the test statistics. Friedman tests are statistically significant for all variables (Tab. 4).

3.2 Mesoscopic analysis: theoretical validation

The simulated market alternates between two regimes: efficient market and speculative bubbles (Fig. 4). These two market regimes arise from the two forecasting strategies available to the agents: respectively, fundamentalism and trend-following (sections 2.2.3 and 2.2.4). The former pushes the price close to the true fundamental value, while the latter amplifies price trends. Individual forecasting strategy choices manage to shape price formation collectively, bottom-up, because of their upstream role in the hierarchical decision-making structure of the agents (section 2.2.1): agents fuel the market with trading orders that embody information about their individual *expectations* (i.e. forecasts) about market behaviour. Once enough traders find it more accurate to employ a given forecasting strategy, their decisions create avalanche effects at the macroscopic level, which push more traders to adapt their strategy to the new market regime. However, the market does not converge to a permanently dominant regime. The environment is constantly evolving, with no safe strategy, in line with seminal work in agent-based modelling [1, 9–11].

The emergence of speculative bubbles is not new in agent-based models of financial markets, but pSYMBA achieves this result by leaving full individual freedom of choice to the agents. For instance, in [33], these dynamics require agents to be biased towards trend-following strategies, rather than contrarian ones (i.e. $P \geq 0$, in their notation). In contrast, in pSYMBA, bubble formation does not require so. Agents are free to choose their own strategy.[2] Moreover, there is no mechanism pushing the agents towards fundamentalism when bubbles arise. In pSYMBA, bubbles do not burst because of a sudden intervention of fundamentalism, but most likely because enough agents stop playing the "greater fool theory" [6], either because the price is too high for them to issue trading orders or because they choose not to buy anymore.

[2]To clarify, in pSYMBA, biases (if activated) are learning biases (e.g. confirmation bias B_+, balanced bias B_0, disconfirmatory bias B_-, Fig. 3): they do not directly apply to strategy preferences, they rather affect how agents learn from the performance of their strategies.

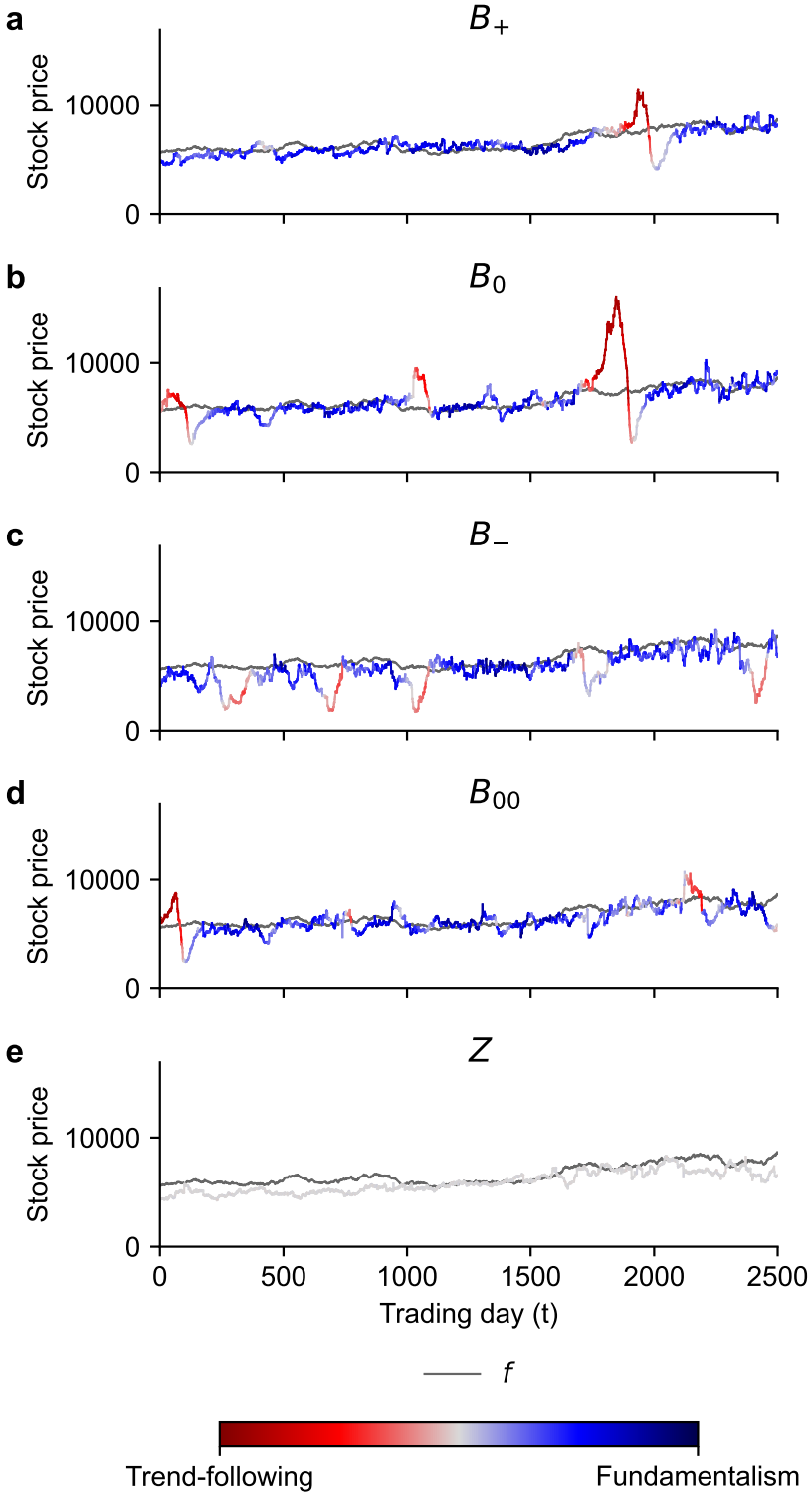

Figure 4: Stock price dynamics: trend-following fuels speculative bubbles. Stock price during the testing phase, across learning conditions (Fig. 3): **a)** confirmation bias (B_+), **b)** balanced bias (B_0); **c)** disconfirmatory bias (B_-); **d)** rationality (B_{00}); **e)** no learning, i.e. zero-intelligence (Z). Price colour represents average strategy choice $\psi_{\mathcal{F},t}$ (eq. 6), i.e. the fraction of agents choosing fundamentalism at a given time step. True fundamental value is shown in dark grey. Mean fundamental value growth is 4.46%/year.

3.3 Microscopic analysis: emergence of strategy preferences aligned with cognitive traits

The model also shows microscopic properties matching trader profiles as expected heuristically: long-term fundamentalists and short-term chartists. We observe that agents develop strategy preferences in line with their individual cognitive traits: more patient and more informed agents become fundamentalists, while more impatient or less informed agents resort to following market trends (Fig. 5). This strategy polarisation is especially enhanced under the confirmation bias condition (Fig. 5a).

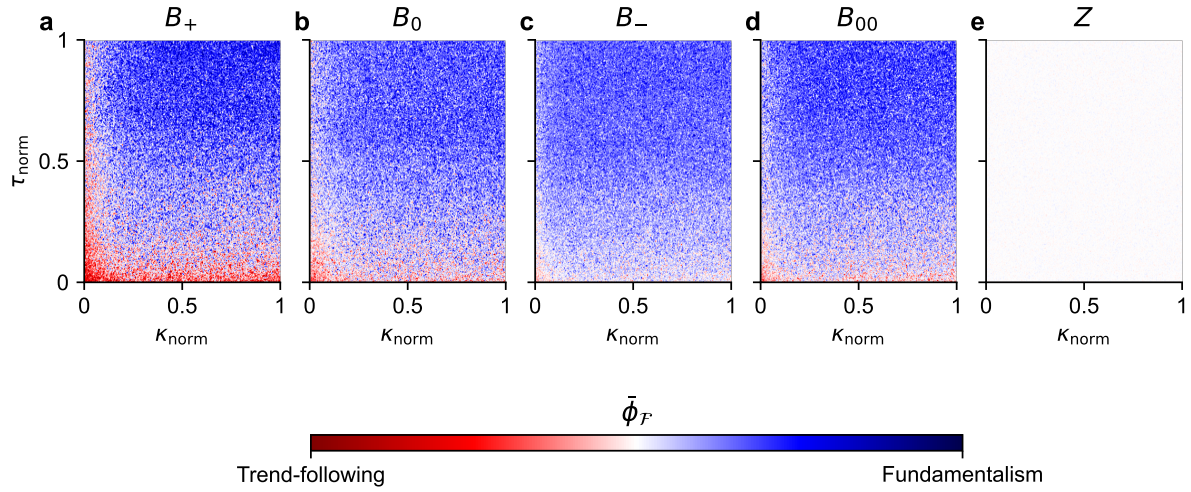

Figure 5: Interpolated forecasting strategy preference $\bar{\phi}_{\mathcal{F},i}$ across learning conditions: news sensitivity κ and patience τ polarise preferences.
a) confirmation bias (B_+), **b)** balanced bias (B_0); **c)** disconfirmatory bias (B_-); **d)** rationality (B_{00}); **d)** zero-intelligence (Z). Learning conditions are described in Methods (Fig. 3). Heatmaps of $\bar{\phi}_{\mathcal{F}}$: forecasting strategy preferences $\phi_{\mathcal{F},i}$ (i.e. the fraction of times that the agent chose fundamentalism over the duration of a testing phase, eq. 7) linearly interpolated across a full dataset of 200 simulations per learning condition. We compute $\phi_{\mathcal{F},i}$ for each agent i, for each testing phase. τ and κ are normalised on the lower and upper bounds of their parameter distributions (Tab. 2).

4 Results

The bias aggregation problem: the impact of confirmation bias

Here, we address the central question of this study: is the aggregated impact of individual confirmation biases accumulating or cancelling out at macroscopic level? When we bias learning in the agents, we find two distinctive effects of confirmation bias: it shrinks market volatility (Fig. 6a), but it exacerbates negative skewness (Fig. 6b). In other words, if agents tend to neglect their bad decisions and stick to their choices, the market becomes more predictable (Fig. 6a), but the relative risk of large price drops increases (Fig. 6b).

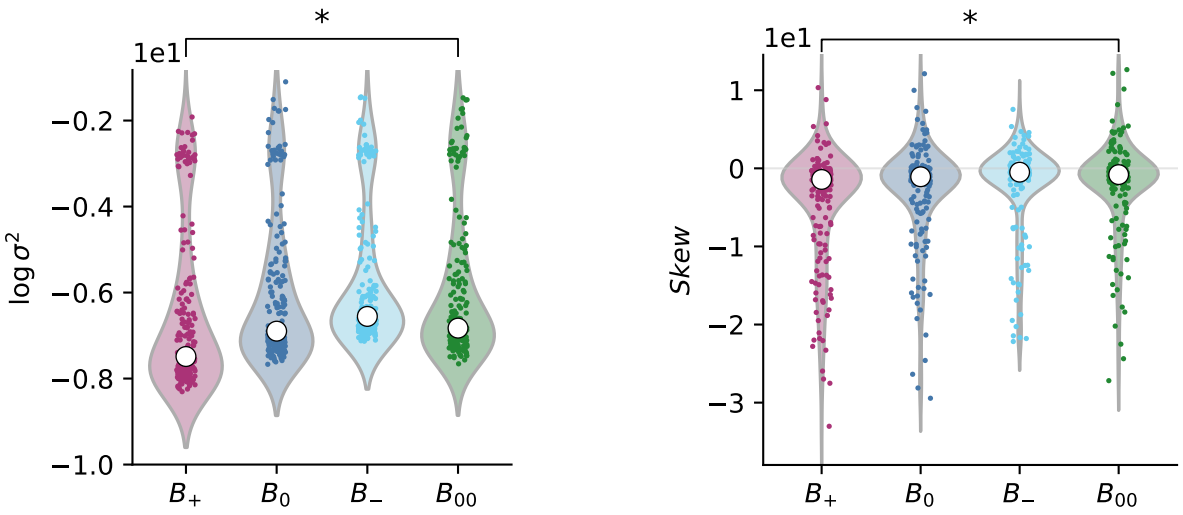

Figure 6: Impact of confirmation bias on volatility and skewness of price returns from simulations.
a) Log-variance of price returns; **b)** Skewness of price returns. Price returns are defined as $r_t = \log P_{t+1} - \log P_t$. Learning conditions are defined in Fig. 3. $^*p < 10^{-5}$ in post-hoc tests (Tab. 7).

We obtain these results from our statistical analysis (section 2.3.1 and Tab. 3). We first run a priori comparison between all learning conditions (Tab. 6), and then post-hoc tests (Tab. 7) contrasting the rationality condition against each biased learning condition.

Variable	$\chi^2(3)$	p
$\log \sigma^2$	93.414	4.05×10^{-20}
$Skew$	45.282	8.06×10^{-10}
$Kurt$	48.342	1.80×10^{-10}

Table 6: Friedman test statistics for a priori comparison between all learning conditions.
Dependent variables are log-variance, skewness and kurtosis. Friedman tests are statistically significant for all variables of interest. P-values are not corrected.

	B_+	B_0	B_-
$\log \sigma^2$	-0.74 $p = 1.52 \times 10^{-6}$ $W = 5933$	-0.09 $p = 0.60$ $W = 9000$	0.21 $p = 0.17$ $W = 8492$
$Skew$	-1.02 $p = 1.85 \times 10^{-6}$ $W = 5964$	-0.29 $p = 0.31$ $W = 8716$	0.06 $p = 1.00$ $W = 9791$
$Kurt$	4.21 $p = 1.00$ $W = 9443$	1.11 $p = 1.00$ $W = 9884$	-5.96 $p = 0.05$ $W = 8102$

Table 7: Impact of learning biases on price returns, with respect to rationality. Post-hoc tests contrast all biased learning conditions to the rationality condition as reference condition (Tab. 3). Columns represent the biased learning conditions: confirmation bias (B_+), balanced bias (B_0), and disconfirmatory bias (B_-), as in Fig. 3. Rows represent dependent variables, i.e. statistical properties of price returns. Each box shows the results from a pairwise Wilcoxon signed-rank test between the column condition and the rationality condition. The top value is the median difference between the two conditions, subtracting the reference condition B_{00} from the column condition. p-values are Bonferroni-corrected by the number of pairwise comparisons (here 3). The bottom value reports the test statistics. Friedman tests are statistically significant for all variables (Tab. 6).

5 Conclusion

This study investigates a *bias aggregation problem* in financial markets, asking whether the collective impact of individual biases on stock price properties cancels out or accumulates. As cognitive bias, we focused on confirmation bias, for its pervasiveness in human cognition and in traders. We address our question with a novel approach, by integrating experimentally-validated behavioural economics models into multi-agent reinforcement learning, to simulate a large number of agents that autonomously drive stock price formation.

The contribution of this study is two-fold. Firstly, it highlights the potential of integrating experimentally-validated cognitive models into financial market simulations. The model is capable of producing emergent properties across levels: from stylised facts of price returns at the macroscopic level, to speculative bubbles at the mesoscopic level, to individual strategy preferences aligned with cognitive traits at the microscopic level. Secondly, this study reveals that individual biases, such as confirmation bias, can shape stock prices. In other words, from a modelling perspective, we collect evidence that the underlying assumptions on traders' cognition do affect macroscopic results.

By combining cognitive science, machine learning and computational finance, this study contributes to the development of the emerging field of computational cognitive finance. Future work could explore other cognitive biases and extend this framework to incorporate more complex forecasting strategies.

References

[1] W Brian Arthur. Inductive reasoning and bounded rationality. *The American economic review*, 84(2):406–411, 1994.

[2] Nicholas Barberis and Wei Xiong. What drives the disposition effect? an analysis of a long-standing preference-based explanation. *the Journal of Finance*, 64(2):751–784, 2009.

[3] Sophie Bavard, Maël Lebreton, Mehdi Khamassi, Giorgio Coricelli, and Stefano Palminteri. Reference-point centering and range-adaptation enhance human reinforcement learning at the cost of irrational preferences. *Nature communications*, 9(1):1–12, 2018.

[4] Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

[5] Colin Camerer. The rationality of prices and volume in experimental markets. *Organizational Behavior and Human Decision Processes*, 51(2):237–272, 1992.

[6] Colin Camerer. Taxi drivers and beauty contests. *Engineering and science*, 60(1):10–19, 1997.

[7] Colin Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton university press, 2011.

[8] Colin Camerer, George Loewenstein, and Matthew Rabin. *Advances in behavioral economics*. Princeton university press, 2004.

[9] Damien Challet and Yi-Cheng Zhang. On the minority game: Analytical and numerical studies. *Physica A: Statistical Mechanics and its applications*, 256(3-4):514–532, 1998.

[10] Damien Challet, Matteo Marsili, and Yi-Cheng Zhang. Stylized facts of financial markets and market crashes in minority games. *Physica A: Statistical Mechanics and its Applications*, 294(3-4):514–524, 2001.

[11] Damien Challet, Matteo Marsili, and Yi-Cheng Zhang. *Minority games: interacting agents in financial markets*. OUP Oxford, 2004.

[12] Gary Charness and Chetan Dave. Confirmation bias with motivated beliefs. *Games and Economic Behavior*, 104:1–23, 2017.

[13] Chu Xin Cheng et al. Confirmation bias in investments. *International Journal of Economics and Finance*, 11(2):50–55, 2019.

[14] Yao-Min Chiang, David Hirshleifer, Yiming Qian, and Ann E Sherman. Do investors learn from experience? evidence from frequent ipo investors. *The Review of Financial Studies*, 24(5):1560–1589, 2011.

[15] James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick. Reinforcement learning and savings behavior. *The Journal of finance*, 64(6):2515–2534, 2009.

[16] Daniel Fonseca Costa, Francisval de Melo Carvalho, Bruno César de Melo Moreira, and José Willer do Prado. Bibliometric analysis on the association between behavioral finance and decision making with cognitive biases such as overconfidence, anchoring effect and confirmation bias. *Scientometrics*, 111:1775–1799, 2017.

[17] Richard M Cyert and Morris H DeGroot. Rational expectations and bayesian analysis. *Journal of Political Economy*, 82(3):521–536, 1974.

[18] John M Darley and Paget H Gross. A hypothesis-confirming bias in labeling effects. *Journal of Personality and Social Psychology*, 44(1):20, 1983.

[19] Chetan Dave and Katherine W Wolfe. On confirmation bias and deviations from bayesian updating. *Retrieved*, 24:2011, 2003.

[20] Peter Dayan and Christopher JCH Watkins. Q-learning. *Machine learning*, 8(3):279–292, 1992.

[21] J Bradford De Long, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann. Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738, 1990.

[22] Jonathan Donier and Jean-Philippe Bouchaud. From walras' auctioneer to continuous time double auctions: A general dynamic theory of supply and demand. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(12):123406, 2016.

[23] Giovanni Dosi, Mauro Napoletano, Andrea Roventini, Joseph E Stiglitz, and Tania Treibich. Rational heuristics? expectations and behaviors in evolving economies with heterogeneous interacting agents. *Economic Inquiry*, 58(3):1487–1516, 2020.

[24] David Easley, Jon Kleinberg, et al. *Networks, crowds, and markets: Reasoning about a highly connected world*, volume 1. Cambridge university press Cambridge, 2010.

[25] Ward Edwards. Conservatism in human information processing. *Formal representation of human judgment*, 1968.

[26] Mahmoud A El-Gamal and David M Grether. Are people bayesian? uncovering behavioral strategies. *Journal of the American statistical Association*, 90(432):1137–1145, 1995.

[27] Euronext. Euronext rule book - book i: Harmonised rules.

[28] Jonathan St BT Evans. *Bias in human reasoning: Causes and consequences.* Lawrence Erlbaum Associates, Inc, 1989.

[29] Eugene F Fama. Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2):383–417, 1970.

[30] J Doyne Farmer. Market force, ecology and evolution. *Industrial and Corporate Change*, 11(5):895–953, 2002.

[31] Cary Frydman and Colin Camerer. The psychology and neuroscience of financial decision making. *Trends in cognitive sciences*, 20(9):661–675, 2016.

[32] Nahuel Garcia, Stefano Palminteri, and Maël Lebreton. The computational origins of confidence biases in reinforcement learning. *PsyArXiv*, 2021.

[33] Irene Giardina and Jean-Philippe Bouchaud. Bubbles, crashes and intermittency in agent based market models. *The European Physical Journal B-Condensed Matter and Complex Systems*, 31:421–437, 2003.

[34] Nicola Giocoli. From wald to savage: homo economicus becomes a bayesian statistician. *Journal of the History of the Behavioral Sciences*, 49(1):63–95, 2013.

[35] William Hart, Dolores Albarracín, Alice H Eagly, Inge Brechan, Matthew J Lindberg, and Lisa Merrill. Feeling validated versus being correct: a meta-analysis of selective exposure to information. *Psychological bulletin*, 135(4):555, 2009.

[36] Cars Hommes and Blake LeBaron. *Computational economics: Heterogeneous agent modeling*. Elsevier, 2018.

[37] Michael C Jensen. The performance of mutual funds in the period 1945-1964. *The Journal of finance*, 23(2):389–416, 1968.

[38] Martin Jones and Robert Sugden. Positive confirmation bias in the acquisition of information. *Theory and Decision*, 50:59–99, 2001.

[39] Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

[40] Daniel Kahneman, Jack L Knetsch, and Richard H Thaler. Experimental tests of the endowment effect and the coase theorem. *Journal of political Economy*, 98(6):1325–1348, 1990.

[41] John Maynard Keynes. *The General Theory of Employment, Interest, and Money*. Palgrave Macmillan Cham, 1936.

[42] Joshua Klayman. Varieties of confirmation bias. *Psychology of learning and motivation*, 32: 385–418, 1995.

[43] Frank Hyneman Knight. *Risk, uncertainty and profit*, volume 31. Houghton Mifflin, 1921.

[44] Finn E Kydland and Edward C Prescott. Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pages 1345–1370, 1982.

[45] Ronald C Lease, Wilbur G Lewellen, and Gary G Schlarbaum. The individual investor: Attributes and attitudes. *The Journal of Finance*, 29(2):413–433, 1974.

[46] Blake LeBaron. Building the santa fe artificial stock market. *Physica A*, 1:20, 2002.

[47] Mael Lebreton, Khalil Bacily, Stefano Palminteri, and Jan B. Engelmann. Contextual influence on confidence judgments in human reinforcement learning. *PLOS Computational Biology*, 15(4):e1006973, 2019. doi: 10.1371/journal.pcbi.1006973.

[48] Daeyeol Lee, Hyojung Seo, and Min Whan Jung. Neural basis of reinforcement learning and decision making. *Annual review of neuroscience*, 35:287–308, 2012.

[49] Germain Lefebvre, Maël Lebreton, Florent Meyniel, Sacha Bourgeois-Gironde, and Stefano Palminteri. Behavioural and neural characterization of optimistic reinforcement learning. *Nature Human Behaviour*, 1(4):1–9, 2017.

[50] Charles G Lord, Lee Ross, and Mark R Lepper. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of personality and social psychology*, 37(11):2098, 1979.

[51] Robert E Lucas Jr. Econometric policy evaluation: A critique. In *Carnegie-Rochester conference series on public policy*, volume 1, pages 19–46. North-Holland, 1976.

[52] Johann Lussange, Alexis Belianin, Sacha Bourgeois-Gironde, and Boris Gutkin. Learning and cognition in financial markets: A paradigm shift for agent-based models. In *Proceedings of SAI Intelligent Systems Conference*, pages 241–255. Springer, 2020.

[53] Johann Lussange, Ivan Lazarevich, Sacha Bourgeois-Gironde, Stefano Palminteri, and Boris Gutkin. Modelling stock markets by multi-agent reinforcement learning. *Computational Economics*, 57:113–147, 2021.

[54] Johann Lussange, Stefano Vrizzi, Sacha Bourgeois-Gironde, Stefano Palminteri, and Boris Gutkin. Stock price formation: Precepts from a multi-agent reinforcement learning model. *Computational Economics*, 61(4):1523–1544, 2023.

[55] Johann Lussange, Stefano Vrizzi, Stefano Palminteri, and Boris Gutkin. Mesoscale effects of trader learning behaviors in financial markets: A multi-agent reinforcement learning study. *Plos one*, 19(4):e0301141, 2024.

[56] Thomas Lux and Michele Marchesi. Scaling and criticality in a stochastic multi-agent model of a financial market. *Nature*, 397(6719):498–500, 1999.

[57] Fred S McChesney. Behavioral economics: Old wine in irrelevant new bottles? *Supreme Court Economic Review*, 21(1):43–76, 2013.

[58] John Stuart Mill. On the definition of political economy; and on the method of investigation proper to it. *London and Westminster Review*, 4(October):120–164, 1836.

[59] Michael P Murray. A drunk and her dog: an illustration of cointegration and error correction. *The American Statistician*, 48(1):37–39, 1994.

[60] John F Muth. Rational expectations and the theory of price movements. *Econometrica: journal of the Econometric Society*, pages 315–335, 1961.

[61] Raymond S Nickerson. Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology*, 2(2):175–220, 1998.

[62] Terrance Odean. Are investors reluctant to realize their losses? *The Journal of finance*, 53 (5):1775–1798, 1998.

[63] Maury FM Osborne. Brownian motion in the stock market. *Operations research*, 7(2): 145–173, 1959.

[64] Stefano Palminteri and Maël Lebreton. Context-dependent outcome encoding in human reinforcement learning. *Current Opinion in Behavioral Sciences*, 41:144–151, 2021. doi: 10.1016/j.cobeha.2021.06.006.

[65] Stefano Palminteri, Mehdi Khamassi, Mateus Joffily, and Giorgio Coricelli. Contextual modulation of value signals in reward and punishment learning. *Nature Communications*, 6:8096, 2015. doi: 10.1038/ncomms9096.

[66] Stefano Palminteri, Germain Lefebvre, E. J. Kilford, and S.-J. Blakemore. Confirmation bias in human reinforcement learning: Evidence from counterfactual feedback processing. *PLOS Computational Biology*, 13(8):e1005684, 2017. doi: 10.1371/journal.pcbi.1005684.

[67] Elise Payzan-LeNestour and Peter Bossaerts. Learning about unstable, publicly unobservable payoffs. *The Review of Financial Studies*, 28(7):1874–1913, 2015.

[68] Charles R Plott. Rational choice in experimental markets. *Journal of Business*, pages S301–S327, 1986.

[69] Sebastien Pouget, Julien Sauvagnat, and Stephane Villeneuve. A mind is a terrible thing to change: confirmatory bias in financial markets. *The Review of Financial Studies*, 30(6): 2066–2109, 2017.

[70] Robert A Rescorla. A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Current research and theory*, pages 64–99, 1972.

[71] Mark Rubinstein. Rational markets: yes or no? the affirmative case. *Financial Analysts Journal*, 57(3):15–29, 2001.

[72] Mohammed Salek, Damien Challet, and Ioane Muni Toke. Price impact in equity auctions: zero, then linear. *arXiv preprint arXiv:2301.05677*, 2023.

[73] W. Schultz, P. Dayan, and P. R. Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997. doi: 10.1126/science.275.5306.1593.

[74] Reinhard Selten. Evolution, learning, and economic behavior. *Games and Economic Behavior*, 3(1):3–24, 1991.

[75] Hersh Shefrin and Meir Statman. The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of finance*, 40(3):777–790, 1985.

[76] Robert J Shiller. Do stock prices move too much to be justified by subsequent changes in dividends? 1981.

[77] Burrhus Frederic Skinner. *The behavior of organisms: An experimental analysis*. BF Skinner Foundation, 2019.

[78] Adam Smith. *The wealth of nations*. 1776.

[79] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[80] H. Thaler, R. and R. Sunstein, C. *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press, 2008.

[81] Richard Thaler. Toward a positive theory of consumer choice. *Journal of economic behavior & organization*, 1(1):39–60, 1980.

[82] Edward L Thorndike. Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, 2(4):i, 1898.

[83] Amos Tversky and Daniel Kahneman. Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *science*, 185 (4157):1124–1131, 1974.

[84] Vanessa Martins Valcanover, Igor Bernardi Sonza, and Wesley Vieira da Silva. Behavioral finance experiments: a recent systematic literature review. *Sage Open*, 10(4): 2158244020969672, 2020.

[85] Martin Weber and Colin Camerer. The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization*, 33(2):167–184, 1998.