Risk-based Adaptive Stock Trading System using Reinforcement Learning

Chai Quek[1], Qi Cao[2*]

***Abstract*—** There is a significant amount of research in applying reinforcement learning (RL) techniques for performing stock trading. Most of these prior works aim to maximize trading profits. However, there is no single type of trader characteristics in stock markets. Different individuals are willing to tolerate differing risk levels in stock trading, which exhibit very different trading behaviors. Firstly, this paper develops a RL agent to model stock trading patterns for maximizing profits, achieved by buying stock near the troughs and selling stock near peaks. Secondly, it is enhanced with a risk-awareness stock trading model that enables tuning the level of risk appetite. This is achieved using a risk-sensitivity parameter and revising the Q-learning algorithm. The proposed risk-awareness stock trading model exhibits a full spectrum of trading behaviors, from risk-averse at one end of the spectrum, to risk-seeking at the other end. Trading behaviors under different risk-sensitivities have been validated against predictions made by the prospect theory, a leading theory for risk-based decision making. It is further validated against human trading behaviors from a user study involving 23 participants. The analysis results illustrate that the proposed risk-awareness trading model can cover wide ranging behavioral characteristics under different risk appetites. Different risk sensitivities are suitable to different stock market conditions. Thirdly, a risk-adaptive stock trading system is developed utilising the RL risk-awareness trading model. It can automatically switch its risk sensitivity profile depending on dynamic stock market conditions. It achieves 4.46% - 10.02% annualized returns on investment for a portfolio of five stocks, achieving superior benchmark returns than trading models using fixed risk-appetites and model with the Moving Average Convergence/Divergence (MACD) algorithm.

***Index Terms*—** Reinforcement learning, Risk-adaptive stock trading, Risk appetite, Risk-awareness agent, Risk sensitivities.

Corresponding author: Qi Cao.

[1] Chai Quek is with School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798. ORCID: 0000-0002-7313-4339. (e-mail: ashcquek@ntu.edu.sg).

[2] Qi Cao is with School of Computing Science, University of Glasgow, Scotland, United Kingdom. ORCID: 0000-0003-3243-5693. (e-mail: qi.cao@glasgow.ac.uk).

1. INTRODUCTION

FINANCIAL stock markets are dynamic and affected by various factors [1][2]. It is highly challenging to be able to accurately model the effective trading behaviour in making judicious decisions to buy, sell or hold a stock. Stock trading systems and trading strategies are interesting research topics with various artificial intelligence (AI) models being reported in the literature [3][4][5]. AI and machine learning models for financial trading applications have been shown to be able to generate profitable trading transactions in the long run [6]. A stock trading system has been reported using trading decisions made by the rough cognitive reasoning networks, where trading opportunities are derived from identified stocks with higher probabilities of rising in the short term i.e., in the next few days [3], based on the momentum investing strategy [7]. Lim et al. [8] introduced a stock trading model combining with deep neural networks (DNN) on trading signal generations to enhance the momentum investing strategy. Another stock trading system was reported to generate trading actions generated by a fuzzy inference engine, based on the detections of technical trading patterns from technical indicators [9]. Tan et al. [10] developed a trading system using an adaptive neuro-fuzzy approach with a decision process to determine stock price cycles through the tuning of the momentum and average periods using reinforcement learning. An intelligent stock trading system was proposed [4] to generate better trading decisions by combining genetic algorithms with rough set analysis, in order to discover optimal or sub-optimal trading rules. Chiang et al. [11] described an adaptive decision support system for stock trading to employ an artificial neural network (ANN) for the prediction of future price movement direction and model creation for individual stock index. A trading system with adaptive control models using deep reinforcement learning and K-lines clustering methods was introduced by [12]. Trading rules are derived from history and current trends using candlesticks in the K-line theory to generalize price movements in a specified time frame. Deng et al. [13] presented their works on stock trading using recurrent DNN for feature learning of trading signals and trading decision making. An enhanced league championship algorithm (LCA) was utilised to develop a rule-based trading system with suitable time of trading stocks being determined in a stock trading rule extraction process [14]. Reinforcement learning (RL) was applied to the LCA network structure to enhance the search and learning capability for extracting better trading rules. A pattern recognition model for stock trading systems using machine learning methods was presented to recognise candlestick patterns to achieve better trading decisions, where the two-day candlestick patterns as well as the three-day candlestick patterns were observed to produce better performance using forecast stock prices one day ahead [15]. Ayala et al. [5] explored a stock trading strategy by combining various machine learning techniques with technical analysis-based strategies to improve trading decisions. Trading performances have been compared among four types of machine learning techniques, out of which ANN and linear model achieved the best results. Trading strategy and decision making using long short-term memory (LSTM) networks were explored in [16].

Stock trading can be modeled as a Markov Decision Process (MDP) [17]. Formalization of sequential decisions can be enabled and performed by MDP. Actions taken in a state affect both immediate rewards in the current state and the subsequent state. MDP is a foundational element of RL, where stock trading can be implemented using RL. Prior works based on reinforcement learning to perform stock trading have been reported. There have been attempts to account for risk in RL based trading systems. Two stock trading strategies using RL and directional change event algorithms are presented to capture stable market states and make better trading actions [18]. Several RL agents have been explored and compared for financial indices trading where RL agent with adaptive learning capability shows better financial returns [19]. Chakole et al. [20] described their stock trading model and dynamic trading strategies using Q-learning method of RL, with two models to form states of the RL environment: historical information clustering model and candlestick model. A RL framework is used in financial data training, where RL agents are trained by financial minute-candle [21]. After such training, stock trading is guided by the RL agents. A RL-based portfolio agent and RL optimization framework are introduced to improve market feedback in response to previous actions by the stock trading agent [22]. The RL agent trained using three hidden layers was developed for a financial trading system to generate trading signals and trading decisions [23]. Chen et al. [24] presented a RL agent for stock trading system with certain trading strategies. The RL was explored in a trading system [25].

Besides, some RL-based algorithms are also utilised to make trading decisions in stock trading systems. Recurrent reinforcement learning (RRL) is a type of adaptive algorithms with previous outputs being fed as a part of the inputs. A RRL algorithm is combined with a particle swarm algorithm to optimize an active portfolio trading system, with notable performance being produced [26]. An equity index trading system is depicted by combining RRL and features extracted from candlesticks patterns [27]. Combinations of various technical indicators in a RRL equity trading system are optimized by a genetic algorithm to improve on the trading profits [28]. The combination of RL and DNN gives rise to the set of deep reinforcement learning (DRL) algorithms [29]. Prior research works show that it is feasible to adopt the DRL in making investment decisions in financial fields, such as stock trading systems [30]. A trading strategy using DRL method and a trading deep Q-network (TDQN) was presented to achieve optimal trading positions for trading activities in financial trading systems, where promising results are reported against benchmark trading strategies [31]. In a developed automated financial trading system, a DRL agent was trained by risk curiosity-driven learning on 504 datasets with a rule-based policy approach [32]. It is able to enhance the quality of actions and trading performance. A DRL framework is trained by learning the environmental representations and selecting features in the financial signals, that makes better action decisions in stock trading [33]. A DRL based day trading system was described using a model-free, off-policy actor-critic method to learn policies in continuous action and state spaces [17]. The trading strategy attempted to execute buy and sell trading on the same day, without holding overnight – essentially an intra contra trade. Park et al. [29] developed a portfolio trading strategy and set of trading actions for multiple financial assets using a DRL agent and deep Q-learning. Trading actions were enabled by mapping functions that can be made for trading directions to each asset such as buying, selling or holding at certain trading size. A DRL-based trading system is trained to reduce the search space and measure the market uncertainty for predicting the number of shares and trading actions more efficiently [34]. Three actor-critic DRL algorithms were combined into an ensemble trading strategy to search the best performing agent within these three DRL according to dynamic market situations [35].

Maximizing profits through trading actions is a common goal of most of the existing stock trading systems. However, the idea that all investors, traders, and portfolio managers in stock market are similar in nature has been debunked in modern behavioral finance research. A common way of segregating them is through their types of risk-appetites, i.e., risk-averse, risk-neutral, or risk-taking [36][37][38]. Usually, investors or traders could be willing to give up some monetary benefits if the possibility of loss can be reduced or avoided. The intensity of such desires or behaviors can be modelled as the term of *risk-sensitivity*, which is to avoid negative outcomes. It results in different trading patterns on investors or traders with different risk sensitivities. A risk-sensitive RL algorithm is reported to integrate risks into RL [39]. A risk-averse trading system is presented by reward adjusted RL [40]. Q-learning algorithms and RL agents with risk awareness and risk-averse are introduced to improve its robustness in trading markets [41]. Market trading agents have been developed using adversarial RL with risk-averse behaviors according to different market conditions [42]. Prior works in the literature attempt to obtain one general model to be applied to all stocks under different market conditions. However, with unpredictable changes in dynamic market conditions, a model handling one stock well may not perform well in other stocks [11].

In this paper, a RL agent is modelled to identify troughs and peaks in time series stock price charts, with suitable actions being taken for buying or selling stocks. By considering the notion of risk appetite in the RL agent, a novel risk-awareness trading agent is proposed which covers the full spectrum of risk-taking behaviour from risk-averse to risk-seeking. Our proposed risk-awareness trading agent leverages on the research of incorporating risk into RL reported in [39]. In the risk-awareness trading agent, the level of risk-appetites can be tuned by a risk-sensitivity parameter. The value of the risk-sensitivity parameter varies in the range [-1, +1], where -1 is associated with the *most risk-averse* level and +1 for the *most risk-seeking* level. The risk-awareness trading agent is validated using two approaches: one as a manifestation of the prospect theory; the other with the user study involving 23 participants. With the novel risk-awareness trading agent being the core module, a novel risk-adaptive stock trading system is then proposed in this paper. It is able to detect the market conditions and automatically switches to a suitable risk-sensitivity profile. The proposed risk-adaptive trading system can be tuned according to the appropriate amount of risk that users (i.e., investors, traders, or portfolio managers) are willing to undertake. It is capable of modeling various trading patterns learned from dynamic market conditions and different risk sensitivities of users.

The main contributions of this paper are as follows.

1) The risk-awareness RL trading agent is proposed to incorporate various risk appetites while maximizing trading profit. It is more effective for the proposed risk-awareness agent with a flexible risk-sensitivity profile which can handle the full spectrum of trading patterns performed by the traders under different risk-sensitivity as well as different stock market conditions, illustrations, multicolor graphs, and flowcharts.

2) The proposed risk-awareness RL trading agent is validated against the predictions on different risk-taking behaviors made by the prospect theory through experiments. The prospect theory explains risk-based decision making in behavioral finance [43]. The experiment results are analyzed in this paper. Another validation for the proposed risk-awareness RL trading agent is conducted by user study experiments with participants to gauge their risk-appetites, where 23 participants are invited to perform stock trading in a simulated environment. Their trading patterns are studied and compared to those of the proposed RL risk-awareness models.

3) The risk-awareness RL trading agent is the core module of the proposed risk-adaptive trading system, that can automatically switch to the appropriate risk-sensitive profiles according to the dynamic and volatile stock market conditions. The investment returns achieved by our proposed trading system have been benchmarked against those with fixed risk appetites trading agents and the trading approach based on the Moving Average Convergence/Divergence (MACD) indicator, which is a popular technical analysis based algorithm used to generate buy and sell signals for stock trading.

The remaining parts of the paper are organized as follows. The proposed risk-awareness stock trading agent with the risk-sensitivity parameter is presented with its trading patterns being analyzed and validated by two approaches in Section 2 Based on the developed risk-awareness trading agent, the risk-adaptive stock trading system is described in Section 3. Section 4 concludes this paper.

2. PROPOSED RISK-AWARENESS STOCK TRADING AGENT

A RL agent is able to interact with its environment at each time step t = 0, 1, 2, 3, ... The environment is represented in the form of a state $s_t \in S$, where S is the set of all available states [44]. With the current state s_t as input, the RL agent takes an action $a_t \in A(s_t)$, where $A(s_t)$ is the set of possible actions being taken in the state s_t. For each action, a reward r_t is received to evaluate the action outcomes, while the state will be moving into the state s_{t+1}. The goal of a RL agent is to determine a policy π and maximize long-term rewards through a series of actions interacting with its environment, represented by a discounted sum shown in Eq. (1).

$$R_t = \sum_{k=0}^{T} \gamma^k r_{t+k+1} \qquad (1)$$

where R_t is the sum of rewards obtained from the time t till to the terminal time T; γ is the discount rate; and r_t is the reward obtained at the time t.

For a policy π, under a particular state $s_t = s$, and taking a particular action $a_t = a$, at the time t, the Q-value (i.e., the value of a state-action pair) is derived by the expected return correspondingly, which is represented in Eq. (2).

$$Q^{\pi}(s,a) = E[R_t | s_t = s, a_t = a] \;\; = E\left[\sum_{k=0}^{T} \gamma^k r_{t+k+1} \middle| s_t = s, a_t = a\right] \qquad (2)$$

where the notation of $Q^{\pi}(s,a)$ is known as the *Q-function*, i.e., the action value function.

2.1 Proposed RL agent for stock trading

We can model stock trading systems as a RL problem. At each time t, the RL agent can take one of two actions in this model. One action is *buying stock*, also known as a *long position*, that is represented by 1. The other is *selling stock*, also known as a *close position*, which is represented by 0. It means the set of possible actions $A(s_t) \in \{0, 1\}$. For simplicity of descriptions, we assume that the agent in this model can only *buy* or *sell* one normalized unit of the stock at a time. This means that at any given transaction, the agent trades a maximum of one unit of the stock.

In our proposed model, at each time t, the state is represented using an instance of historical stock price changes. The fractional change Δ_t in stock price at time t is given in Eq. (3).

$$\Delta_t = (p_t - p_{t-1})/p_{t-1} \qquad (3)$$

where p_t is the stock price at time t. The state vector at time t is updated accordingly, as shown in Eq. (4).

$$s_t = [\Delta_{t-h+1} \quad \Delta_{t-h+2} \quad \Delta_{t-h+3} \quad \dots \quad \Delta_t \quad pos_t] \qquad (4)$$

where *h* represents the number of historical price changes. The notation *pos*$_t$ represents if the agent currently holds a stock. The value of *pos*$_t$ = 1 if the agent owns a stock, else the value of *pos*$_t$ = 0. The state vector *s*$_t$ captures the price changes in the past number of *h* trading days. It also captures the information on whether the agent currently holds a stock. Instead of the traditional approach of using the absolute prices of the stocks, our proposed model is built with changes in the stock prices only as parts of the states. This proposed approach has two major advantages as follows.

- Since the agent learns from the trends in price changes, it is able to react to situations that it has not directly encountered previously.
- The learning is general in nature and can be applied to any stock. It means even if the training of the RL agent is based on one stock, it is able to extend this ability to trade on other stocks.

As a type of neural networks, Q-network is used as a function approximator for the Q-function. The Q-network takes a state *s*$_t$ as input and derives an output vector representing the Q-values for possible actions. In the proposed model, there are two possible actions in *A*(*s*$_t$): *buy* and *sell*. At each time step *t*, the agent takes an action *a*$_t$ ϵ *A*(*s*$_t$) with the highest Q-value under the policy, as shown in Eq. (5).

$$a_t = \pi(s_t) = arg \max_{a \epsilon A(s_t)} Q(s_t, a) \qquad (5)$$

The optimal trading strategy is to perform *buy* transactions when stock prices start rising (i.e., at *troughs*) and *sell* when stock prices start falling (i.e., at *peaks*). The rewards are structured in a way that the RL agent is capable of learning such behaviors. Stock trading also involves commission costs incurred for every transaction when stocks are bought or sold. These commission costs are incurred by the commission rates charged according to the dollar amounts that are being traded. We incorporate commission rate per transaction as a parameter δ in our model to avoid excessive trading.

At any time step *t*, the action *a*$_t$ determines the reward *r*$_{t+1}$ received in the next time step *t* + 1, as shown as follows.

- When *a*$_t$ = 1 for a *buy* action:

➢ if the value of *pos*$_t$ = 0 (i.e., not holding the stock), a positive reward is received when the stock prices are rising. While a negative reward is received when the stock prices are falling.

➢ if the value of pos_t = 1 (i.e., holding the stock), then the *buy* action is equivalent to doing nothing since the agent can only hold a maximum one unit of the stock. As such, the reward received is zero.

● When a_t = 0 for a sell action:

➢ if the value of pos_t = 1 (i.e., holding the stock), a positive reward is received when the stock prices are falling. While a negative reward is received when the prices are rising.

➢ if the value of pos_t = 0 (i.e., not holding the stock), then the *sell* action is equivalent to doing nothing. Hence, the reward is zero.

It is accommodated in the reward structure shown in Eq. (6).

$$r_{t+1} = (a_t - pos_t)(p_{t+1} - p_t) - |a_t - pos_t|\delta p_t \qquad (6)$$

where r_t represents the reward received at time t. The basic idea behind the Q-learning algorithms is to iteratively estimate the Q-value using the Bellman equation [44]. The temporal difference TD_t at time t for a single step of Q-learning is defined in Eq. (7).

$$TD_t = \ r_t + \gamma \max_{a\prime \in A(s')} Q(s_{t+1}, a') - Q(s_t, a_t) \qquad (7)$$

where r_t is the reward received at time step t; γ is the discount rate; $Q(s_t, a_t)$ is the original value of the network. With the value of the temporal difference TD_t, the recurrent step to update the value of $Q(s_t, a_t)$ for the Q-learning algorithms is shown in Eq. (8).

$$Q(s_t, a_t) \leftarrow \ Q(s_t, a_t) + \ \alpha\, TD_t \qquad (8)$$

where α is the learning rate, multiplying with the value of TD_t.

When a neural network is used to approximate the Q-function, we train it by minimizing an appropriate loss function. We define the loss function for our model as shown in Eq. (9), to update the weights of the network.

$$Loss = \frac{1}{2}[r_{t+1} + \gamma \max_{a\prime} Q(s_{t+1}, a') - Q(s_t, a_t)]^2 \qquad (9)$$

where the term of $r_{t+1} + \gamma \max_{a\prime} Q(s_{t+1}, a')$ is the target value as the output from the network. We use stochastic gradient descent with backpropagation to adjust the weights of the network. In order to achieve stable training, two techniques are used, i.e., a separate target network and experience replay, that have been reported to drastically improve the training of a Q-network [45]. We improve the training of our network by using the method of prioritized experience replay [46]. It enhances the traditional experience replay technique by prioritizing some events over others, where the probability of selecting an experience is proportional to its temporal difference value. Besides, we also use the technique of double-Q learning reported in [47], to correct the overestimation of Q-values by the networks. The RL model developed in this sub-section attempts to maximize the profit judiciously derived from its stock trading actions. However, individuals with different risk appetites often perform trading differently under stock markets as well as market conditions. For example, risk-averse individuals are willing to forego some potential profit to avoid a high probability of losing money.

A general framework is reported previously to incorporate risks into RL algorithms by directly modifying the temporal difference [39]. In this paper, we propose an approach to develop a risk-awareness trading agent by adopting this framework. This approach can be readily applied to supplement the stock trading model developed in this sub-section. As such, the proposed risk-awareness trading model is capable of integrating the considerations on full spectrum of risks factors and risk sensitivities, which can model trading behaviors of different traders, investors or portfolio managers.

2.2 Our methodology of risk-awareness stock trading agent

In our risk-awareness approach, a parameter $k \epsilon$ [-1, 1] is defined as the scalar parameter that controls *risk-sensitivity* of the agent, whose value is elaborated as follows.

- -1 ≤ k < 0: leads to risk aversion.
- k = 0: leads to risk neutrality.
- 0 < k ≤ 1: leads to risk seeking.

In this risk-awareness approach, the temporal difference *TD* is transformed using the parameter k by the transformation function defined in Eq. (10).

$$X: TD_t \mapsto \begin{cases} (1+k)TD_t & if\ TD_t \geq 0 \\ (1-k)TD_t & otherwise \end{cases} \qquad (10)$$

Hence, we revise the value of $Q(s_t, a_t)$ for the Q-learning algorithms in Eq. (8) into the Q-learning rule, by replacing the term $\alpha\ TD_t$ by the term $\alpha X(TD_t)$, as shown in Eq. (11).

$$Q(s_t, a_t) \leftarrow\ Q(s_t, a_t) +\ \alpha X(TD_t) \qquad (11)$$

It allows us to set the risk-sensitivity on a sliding scale value in the range from -1 to 1. If the agent is risk-averse (i.e., $-1 \leq k < 0$), the temporal difference value is amplified when its value is negative. The temporal difference value is reduced when its value is positive. It means that the risk-averse agent learns more from its losses than its gains. It is more sensitive to penalties than rewards.

The opposite is true for a risk-seeking agent (i.e., $0 < k \leq 1$) which learns more from its gains than losses. For the risk-seeking agent, the temporal difference value is amplified when its value is positive, according to Eq. (10). While the temporal difference value is reduced when its value is negative.

A risk-neutral agent (i.e., $k = 0$) places equal weight on losses and gains and learns equally from them. In this case, the risk-awareness model is equivalent to the model described in Sub-section 2.1. Hence, the model developed in Sub-section 2.1 is a special case of this proposed risk-awareness model.

The analysis of the two extreme values of the risk-sensitivity parameter k, i.e., $k = -1$ and $k = 1$, are as follows.

- **When k = -1:** This is the case of extreme risk-aversion. The agent assumes that the worst-case outcome occurs in any situation. As such, it only learns from its losses and disregards any gains.
- **When k = 1:** This is the case of extreme risk-seeking. The agent assumes that the best-case outcome occurs in any situation. It only learns from its gains and ignores the losses.

Transforming temporal differences as done in Eq. (10) using the parameter k to accommodate risk is supported by research in neuroscience. Different models of temporal difference learning have been compared according to the way that a temporal difference is processed by human brains using functional magnetic resonance imaging (fMRI). It has been found that transforming temporal difference according to risk-sensitivity best represents the way that risk-based RL processing occurs in human brains [48][49].

2.3 Experimental results of proposed risk-awareness agent

Experiments with different risk profiles were conducted on different types of stock market conditions. The proposed RL risk-awareness agent is simulated on the following stocks and market conditions.

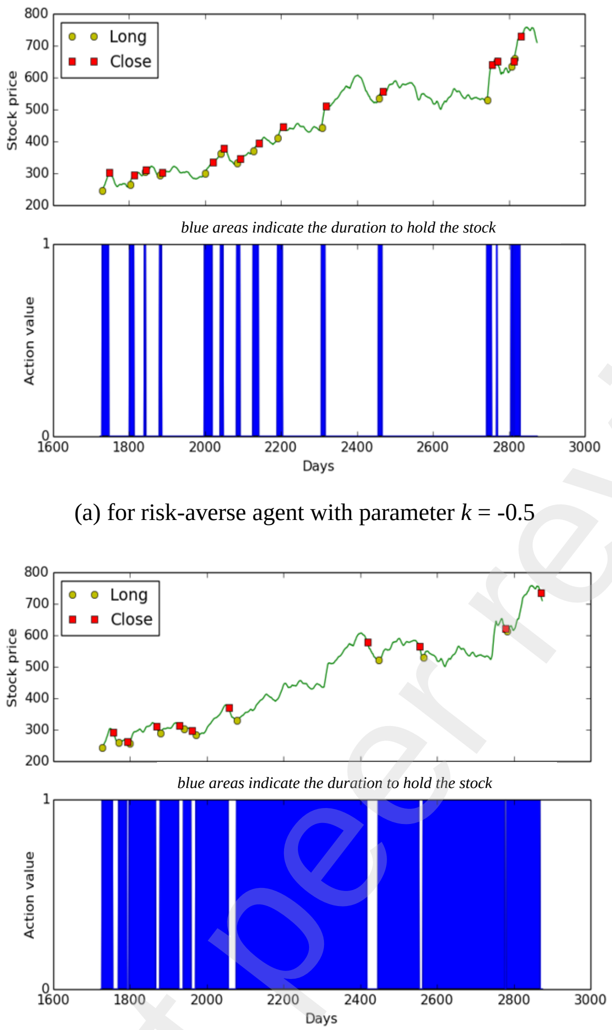
- **Bullish market** – when stock prices are generally rising. The historical prices of Google stock (GOOG) were used for the experiments during 27th June 2011 to 22nd January 2016.
- **Bearish market** – when stock prices are generally falling. The historical prices of Bank

of America Corp stock (BAC) were employed for the experiments during 15th March 2010 to 8th December 2011.

- **Catastrophic market** – when stock prices are rapidly crashing to lose near 100% of its value with no hope of recovery in the near future. The historical prices of American International Group stock (AIG) were used in the duration of 31st July 2007 to 7th October 2008.
- **Volatile market** – when there is high volatility in the stock market, where prices are rapidly changing up and down. The iShares MSCI Hong Kong Index (EWH) was utilized during 4th September 2007 to 27th April 2010.

The experiments were performed under different market conditions by configuring the proposed risk-awareness agent to be risk-averse and risk-seeking types. To simplify the descriptions and analysis purpose of the experiments, the value of the risk-sensitivity parameter k is set to be the mean values representing the risk aversion and risk seeking types respectively, i.e., k = -0.5 is set for the risk aversion agent; and k = 0.5 is set for the risk seeking agent.

It is observed from Fig. 1, the RL agent is able to successfully identify troughs and peaks in the stock data time series and accordingly undertake the appropriate actions. In the bullish market experiment, the trading patterns on the GOOG stock for the configured risk-averse agent are shown in Fig. 1(*a*). Its upper half figure shows that suitable *buying* (i.e., long position) or *selling* (i.e., close position) actions are performed near troughs and peaks. The blue regions in the lower half figure are represented the durations when the risk-averse agent holds this stock after each buying action and before selling it. The trading patterns on the same stock for the configured risk-seeking agent are shown in Fig. 1(*b*). Similarly, its upper half figure shows that suitable *buying* or *selling* actions are identified near troughs and peaks. The blue regions in the lower half figure are represented the durations the risk-seeking agent holds this stock after each buying action. Comparing Fig. 1(*a*) and 1(*b*), it is observed that the risk-seeking agent performs a smaller number of trading actions while holding the stock for much longer period of durations. The risk-seeking agent rides on the price rising trends under the bullish market to gain higher trading profit than that of the risk-averse agent.



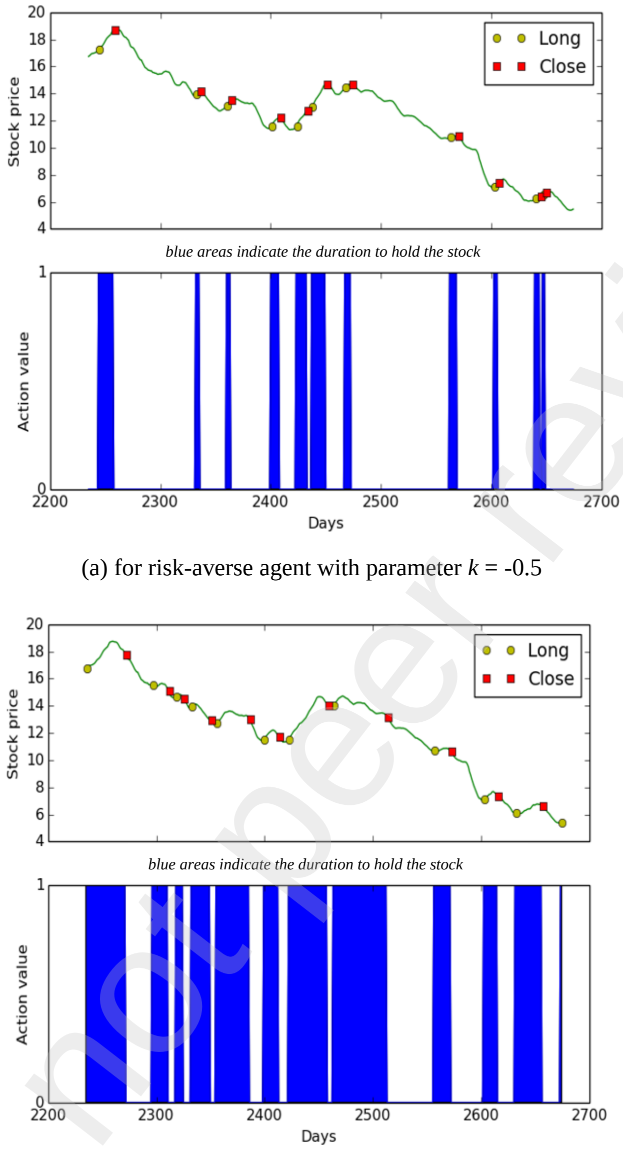
(a) for risk-averse agent with parameter *k* = -0.5

(b) for risk-seeking agent with parameter *k* = 0.5

Fig. 1. Trading patterns on GOOG stock of proposed risk-awareness agents under bullish market.

The experiment results in Fig. 1 clearly illustrate the different trading behaviors and patterns caused by the different risk-sensitivity agents. The differences of trading behaviors and patterns will become even larger, if the value of the risk-sensitive parameter k shifts towards two extreme values with *k* = -1 and *k* = 1.

For the experiment under the bearish market, the trading patterns on the BAC stock for the configured risk-averse agent are shown in Fig. 2(*a*). While the trading patterns on the same stock for the configured risk-seeking agent are shown in Fig. 2(*b*). The upper half figures indicate the *buying* or *selling* actions to be performed in the time series data of the BAC stock. The blue regions in the lower half figures are represented the durations when the risk-awareness agents hold the stock after each buying action and before selling it. Different trading behaviors between the risk-averse agent and the risk-seeking agent are obviously observed. It is shown from Fig. 3 that the risk-averse agent holds the stock for shorter period of durations to reduce the potential losses on the price falling trends under the bearish market, compared to that of the risk-seeking agent.



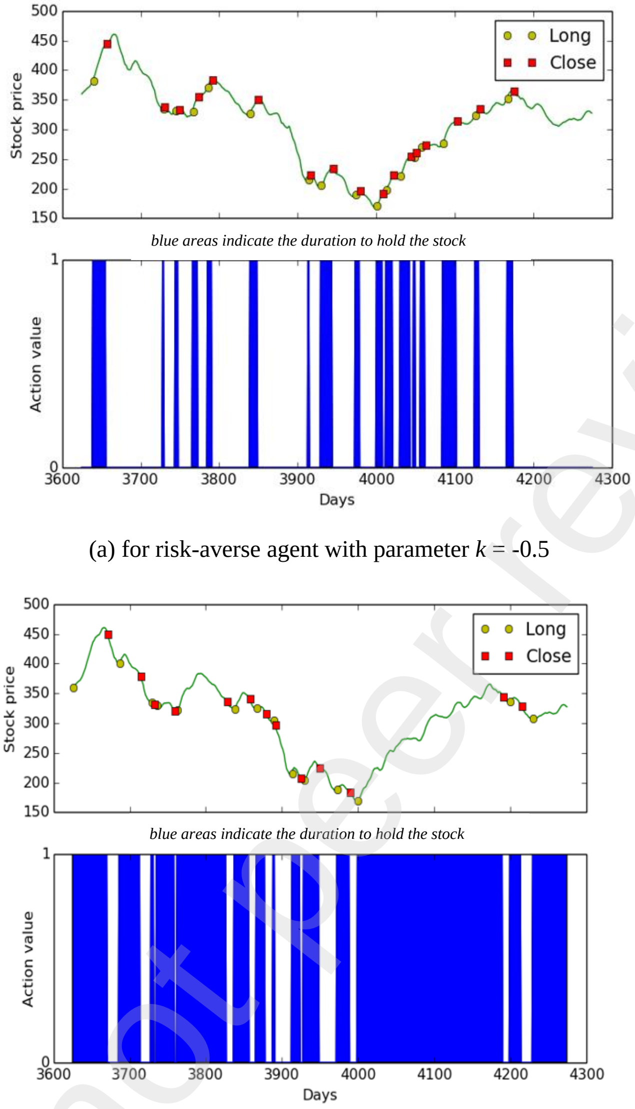
(a) for risk-averse agent with parameter *k* = -0.5

(b) for risk-seeking agent with parameter *k* = 0.5

Fig. 2. Trading patterns on BAC stock of proposed risk-awareness agents under bearish market.

In the experiment under the volatile market, the stock trading patterns on the EWH index for the configured risk-averse agent are shown in Fig. 3(*a*), while those of the risk-seeking agent are shown in Fig. 3(*b*).

Besides the graph illustrations on differences of trading patterns under various market conditions shown in Fig. 1 – Fig. 3, more detailed quantitative analyses have been conducted with five different values of the risk-sensitivity parameter *k* (i.e., *k* = -0.5, -0.2, 0, 0.2, 0.5). It is to compare and show the trends of trading behaviors in the spectrum of risk profiles, including risk-averse, risk neutrality and risk-seeking under various market conditions.



(a) for risk-averse agent with parameter *k* = -0.5

(b) for risk-seeking agent with parameter *k* = 0.5

Fig. 3. Trading patterns on EWH index of proposed risk-awareness agents under volatile market.

The following parameters will be compared and analyzed across different risk profiles in the experiments for the proposed RL risk-awareness agent.

- **Average holding days** (*Average_holding*): The average number of days that the agent holds a stock after buying it. It means the regions where the Q-value of action '*buy stock*' is greater than Q-value of action '*sell stock*'. This is a measure of how quickly the agent sells the stock after buying it.
- **Percentage of holding period** (*Percentage_holding*): The percentage of the total number of days of holding a stock after buying it, over the total number of trading days in the time frame under analysis.
- **Annualized return on investment**: It is the investment return achieved under different

types of market conditions by the risk-awareness agent with different risk sensitivities. The returns are benchmarked against the trading algorithm of Moving Average Convergence/Divergence (MACD) [50][51].

The number of *average holding days* tells us how quickly a risk-awareness agent sells its held stock (or closes its position) after the buying actions. Let it be represented by the parameter *Average_holding*. The experiment results for the *average holding days* of the agent with five types of risk profiles under various market conditions are shown in Table 1. The value trends of the parameter *Average_holding* for these risk-awareness agents can be observed and identified as shown in Eq. (12).

$$Average\_holding_{bullish} > Average\_holding_{bearish} > Average\_holding_{catastrophic} \qquad (12)$$

The trends shown in Eq. (12) are satisfactory because in bearish and catastrophic markets, traders or investors need to sell their holding stocks quickly in the wake of falling prices to avoid losses.

Table 1. number of *average holding days* of different risk-awareness agents under various market conditions

	***k* = -0.5**	***k* = -0.2**	***k* = 0**	***k* = 0.2**	***k* = 0.5**
Bullish	10.93	23.98	55,47	67.14	102.20
Bearish	7.09	10.05	17.92	18.27	23.82
Catastrophic	0.00	3.66	5.79	12.90	15.55
Volatile	8.23	14.60	24.26	29.91	35.15

A more intriguing trend is seen across different values of the parameter *k*. It can be observed from Table 1 that as the value of parameter *k* increases, the value of *Average_holding* also increases. This trend is visible under all stock market conditions. It is expected because a risk-averse agent tries to play it safe and cashes out its rewards quickly, as fearing it might lose its existing profits. On the other hand, a risk-seeking agent does not sell its stocks, even if stock prices start to fall because it is less affected by a possibility of loss. As such, the number of average holding days that a stock is held ought to be positively correlated with risk-seeking tendency. This is clearly visible in the trading patterns in the bullish market as shown in Fig. 2. The risk-averse agent repeatedly sells the stock to cash in the small positive profits whereas the risk-seeking agent keeps holding on to the stock. Furthermore, it can be observed that in a bearish market as shown in Fig. 3, the risk-averse agent sells its stock almost immediately whereas the risk-seeking agent waits longer before selling its stock and takes a risk that prices might rise in the future.

The values of *percentage of holding period* for the risk-awareness agent with different risk profiles under various market conditions are shown in Table 2. The parameter of *percentage of holding period* represents how willing a particular agent is to keep money invested in a stock and hold the stock. Let this parameter be represented as *Percentage_holding.* It is observed that the value trends of *Percentage_holding* for all these risk-awareness agents under various market conditions are shown in Eq. (13).

$$Percentage\_holding_{bullish} > Percentage\_holding_{bearish} > Percentage\_holding_{catastrophic} \quad (13)$$

The value trends shown in Eq. (13) indicate that the proposed risk-awareness agent works properly because it is more profitable to keep money invested in bullish markets than bearish ones.

Table 2. *Percentage of holding period* by different risk-awareness agents under various market conditions

	***k* = -0.5**	***k* = -0.2**	***k* = 0**	***k* = 0.2**	***k* = 0.5**
Bullish	14.26%	66.60%	81.73%	82.00%	88.86%
Bearish	12.72%	42.21%	41.13%	49.54%	60.00%
Catastrophic	0.00%	11.01%	36.63%	46.67%	47.33%
Volatile	21.53%	56.15%	61.69%	63.07%	77.38%

It is also observed from Table 2 that the values of *Percentage_holding* increase as the value of the parameter *k* becomes larger. This shows that the willingness to hold a stock increases as the risk-seeking tendency of the agent increases. It is observed to be true across all stock market conditions. This is expected because a risk-seeking agent should be more willing to invest money in a stock than a risk-averse agent. The risk-averse agent will try to be more certain on its bet. It will not invest money in a lot of uncertain situations where the risk-seeking agent might do.

Table 3. Annualized returns on investment by different risk-awareness agents in various conditions and MACD algorithm

	***k* = -0.5**	***k* = -0.2**	***k* = 0**	***k* = 0.2**	***k* = 0.5**	***MACD algo***
Bullish	8.73%	15.16 %	32.42%	28.88%	52.99%	15.07%
Bearish	-6.87%	-9.13%	-13.42	- 20.83%	- 32.54%	-31.47%

Catastrophic	0.00%	-14.38%	-32.48	-63.25%	-85.84%	-79.85%
Volatile	-2.53%	3.84%	1.88%	8.83%	-12.10%	-9.46%

As a result, the values of *Percentage_holding* ought to be positively correlated with risk-seeking tendencies. It is quite clearly visible of the trading patterns in the bear market, as shown in Fig. 3. The risk-seeking agent is willing to hold the stocks even if the prices are falling. While the risk-averse agent stays clear and only invests when signs of upward price trends are present. It is observed starkly from Table 1 and Table 2 that the risk-averse agent with k = -0.5 refuses to enter the catastrophe market at all in times, as the values of *Percentage_holding* = 0%.

The annualized returns on investment from stock trading achieved by the risk-awareness agents under different risk profiles as well as the benchmark trading model with the MACD algorithm are shown in Table 3. It is observed from Table 3 that the returns achieved by the proposed RL risk-awareness agent under different risk profiles are generally higher than those derived from the MACD algorithm. It shows that the risk-awareness agent is capable of successfully achieving a profit maximizing trading strategy, by performing trading actions near peaks and troughs correctly. In addition, it is capable of accounting for different risk sensitivities and risk profiles. Furthermore, it is also evident that different levels of tunable risks are suitable to different stock market conditions. Being risk-seeking leads to higher returns during a bullish run and being risk-averse leads to lower losses in bearish and catastrophic markets. The proposed risk-awareness agent serves as the core module to develop a risk-adaptive trading system to be described in Section 3.

2.4 Validation of trading behavior using prospect theory

In order to validate the trading behavior of the proposed RL risk-awareness agents, the prospect theory (PT) is employed, that is an influential theory in behavioral finances. It explains successfully many real-world decisions making under risk conditions [43][52]. It accounts for the shortcomings of the expected utility theory (EUT) and the empirical evidence that shows violation of the EUT.

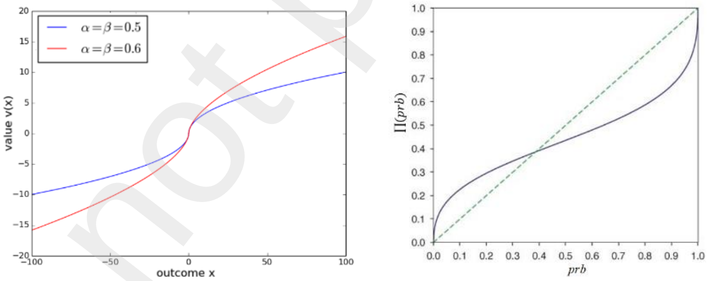
The EUT's utility function is replaced by the PT with a value function v(). It is defined by relative gains and losses from a reference value, instead of absolute values of outcomes. The objective probabilities *prb* of the EUT are also transformed into subjective probability weighting function Π(*prb*) for the PT. Hence, the value V of a prospect that pays additional outcome $\$x$ with probabilities *prb* is shown in Eq. (14).

$$V(x, prb) = \prod(prb)\, v(x) \qquad (14)$$

If the value of the outcome *x* is positive, it represents the region of gains. While a negative value of the outcome *x* represents the region of losses. The value function *v(x)* is defined in Eq. (15).

$$v(x) = \begin{cases} x^{\alpha} & x \geq 0 \\ -\lambda(-x)^{\beta} & x < 0 \end{cases} \qquad (15)$$

where parameters of α and β control the risk-sensitivity of a trader or investor. The parameter α represents the risk-sensitivity in the case of profit gains with rising stock prices. While the parameter β represents the risk-sensitivity in the case of losses with falling stock prices. The risk-seeking tendencies increase as the values of parameters α and β increase. The parameter λ represents *loss aversion*, which is a measure of how much a trader dislikes losses compared to gains. The plots of the value function *v*(*x*) are changed according to the changes in values of parameters α and β, as shown in Fig. 4(*a*). The plots of the value function *v*(*x*) are *concave* in the region of gains, and *convex* in the region of losses. The plots of the probability weighting function Π() are *concave* for small probabilities *prb* and *convex* for large probabilities as shown in Figure 4(*b*).



(a) Plots of value function with different risk-sensitivities

(b) Plots of probability weighting function

Fig. 4. Trading patterns on EWH index of proposed risk-awareness agents under volatile market.

PT can be used to predict the behavior of traders with different level of risk-sensitivities. Assume a trader bought a stock at the price *$X*. Its price has now risen to $(*X* + δ_1) on this day. On the next day, there are two possible scenarios on the stock price. The stock price can further increase to $(X + δ_1 + δ_2) with the probability *prb* or fall back to $X with the probability (1 - *prb*). According to the PT, the value of selling the stock on this day is derived in Eq. (16).

$$v_{sell} = v(\delta_1) = {\delta_1}^{\alpha} \qquad (16)$$

where the parameter α is the risk-sensitivity of the trader. Similarly, the value of holding the stock (i.e., not selling) on this day is given in Eq. (17).

$$v_{hold} = prb \times v(\delta_1 + \delta_2) + (1 - prb)v(0) = \ prb \times v(\delta_1 + \delta_2) = prb \times (\delta_1 + \delta_2)^{\alpha} \qquad (17)$$

The traders have an incentive to hold their stock and not sell if values of $v_{hold} > v_{sell}$, according to Eq. (16) and Eq. (17). The implication of this condition is derived as shown in Eq. (18).

$$v_{hold} > v_{sell} \quad \rightarrow \qquad prb \times (\delta_1 + \delta_2)^{\alpha} > \delta_1{}^{\alpha} \quad \rightarrow \quad \alpha > -\frac{\log{(prb)}}{\log{(1+\frac{\delta_2}{\delta_1})}} \qquad (18)$$

It means that for the given values of *prb*, δ_1, and δ_2, the parameter α has to be above a threshold value in order for the traders to hold their stock shown in Eq. (18). This implies that traders with a higher value of parameter α have the greater incentive to hold the stocks. While a lower value of parameter α implies a greater incentive to sell the stocks for the given values of *prb*, δ_1, and δ_2. It shows that the PT predicts that risk-seeking individuals (i.e., higher value of parameter α) are more likely to hold on to their stocks than risk-averse individuals (i.e., lower value of parameter α).

A similar analysis can be carried out in the case of loss when the stock price falls. Risk-seeking individuals (i.e., higher value of parameter β) are more likely to hold on to the stock as compared to risk-averse individuals (i.e., lower value of parameter β) who will sell it more quickly.

The observations and analyses match with the trading behavior of the proposed RL risk-awareness agents described in Sub-section 2.3, where a risk-seeking RL agent holds its stock for longer than a risk-averse RL agent. Hence, the behavior predicted by the PT is in agreement with the behavior exhibited by the proposed risk-awareness RL agent. It is an encouraging validation of the risk-awareness RL agent by a leading theory about risk-based decision making in behavioral finance.

2.5. Validation of Trading Behavior by User Study

In order to further validate the behavior of the proposed risk-awareness RL agent, another user study has been conducted to assess the risk-appetite of potential users by asking them to perform stock trading in a simulated environment. Their trading patterns are compared to those of the proposed risk-awareness agent with different level of risk-sensitivity profiles.

A task named Balloon Analogue Risk Task (BART) is reported in prior studies to measure the risk appetite of users in a controlled setting [53]. Evidence has been reported about the correlations between users' performances on the BART and risk-taking tendencies in the real world [53][54]. Hence, the BART is employed in the user study for gauging the risk appetite of users.

There were 23 university students at their Year 3 – 4 study from Nanyang Technological University (NTU), Singapore participating in the experiment. The experiment was administered online using a web application developed by our research team. Each participant is provided a link to this web application and asked to complete the tasks. The experiment comprised of two steps as follow.

1) The risk appetite of each participant is measured using the BART. It is to classify the participant according to their risk profiles as risk-averse or risk-seeking on the basis of their performance in the BART experiment.

2) Each participant was asked to perform trading on a bullish stock, a bearish stock, and a volatile stock in a simulated environment. Their results are used to compare against those derived from the proposed risk-awareness RL agent.

Trading patterns of the participants were recorded, which will be analyzed again according to two parameters introduced in Sub-section 2.3, i.e., average holding days (*Average_holding*) and percentage of holding period (*Percentage_holding*).

Step 1 - BART task by the participants

On the BART, detailed instructions are provided to the participants before they attempt the task. In this research, we adopt the simplified version of the BART reported by van Ravenzwaaij et al. [55], which has been shown to achieve better isolations of the risk tendencies of participants. According to [55], the risk-taking tendency defined as notation γ is shown in Eq. (19).

$$\gamma' = -\omega \log(1 - pum) \qquad (19)$$

where the notation ω is the average number of pumps; the notation *pum* is the probability of the balloon bursting on every pump. A higher value of γ' signifies a greater risk-taking tendency. The BART task has the following characteristics and user options.

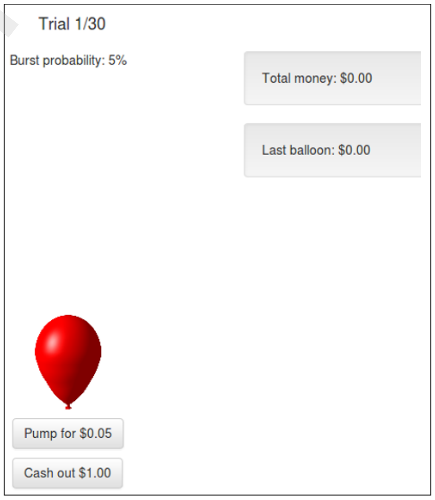
- **Pump** – As one user option, the participant can pump each balloon to increase its size and add some money to the temporary bank account. However, every time the balloon is pumped, there is probability that the balloon will burst. Then, all the money accumulated by the participant in the temporary bank account will be lost.
- The parameter *pum*, i.e., probability of the balloon bursting on every pump, is a constant for the entire task. It is set to *pum* = 5% in the experiment.
- **Cash out –** The other user option is to play it safe without or further pumping the balloon and cash out the money from the temporary bank account to his/her permanent bank account.
- The reward for every pump is determined such that the expected return of each pump is zero if the balloon burst.

In the BART experiment, each participant is presented with total 30 balloons and one at a time. The participants can decide to take actions to **pump** and **cash out**. A trial ends when either the balloon bursts or the participant chooses to cash out the money. An example of visual description of the task is shown in Fig. 5(*a*). Each participant is asked to maximize the money in his/her permanent bank account over 30 trials. The total number of pumps and money are recorded for all participants in the BART experiment.

This BART task captures a defining trait of risk-taking tendencies in the stock market – balancing possible reward against some chance of losses. The derived values of parameter γ' in Eq. (19) are the measure of risk-appetite for the participants derived from the BART. It is best interpreted as a relative measure of risk. It tells us whether a participant is more or less risk-seeking than another participant. As a result, all of the participants are ranked in decreasing order by their derived values of parameter γ', with the median value of parameter γ' is accordingly identified. Those participants with their γ' values above the median value are classified as risk-seeking type. Those with their γ' values below the median value are classified as risk-averse type.

After the BART experiment, 15 out of the 23 participants are classified as risk-averse, and the remaining 8 participants are classified as risk-seeking. The average number of pumps and average value of parameter γ' are shown as follows.

- For all risk-averse participants: average number of pumps = 8.56; average value of parameter γ' = 0.19.
- For all risk-seeking participants: average number of pumps = 17.88; average value of parameter γ' = 0.39.



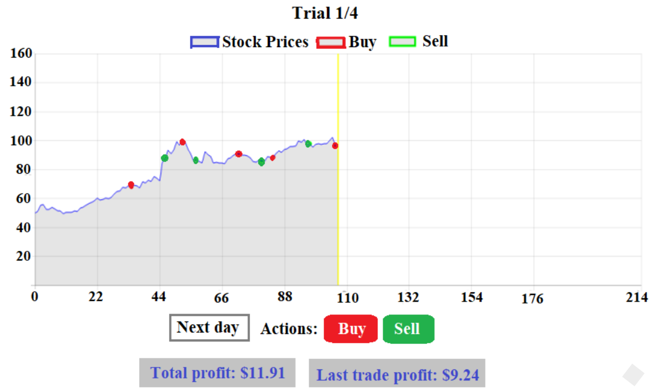
Trial 1/30

Burst probability: 5%

Total money: $0.00

Last balloon: $0.00

(a) Step 1 - Example view of the BART task



Total profit: $11.91

Last trade profit: $9.24

(b) Step 2 - Example view of simulated stock trading task

Fig. 5. Visual descriptions of the two steps of user study experiment.

Step 2 - Simulated stock trading task by the participants

Once the risk appetites of the participants have been estimated by the BART experiment, they were tasked to perform 180 days of stock trading for three stocks in a simulated environment. Detailed instructions are provided to the participants before they begin this task.

This task is designed to simulate the RL agent's environment as closely as possible. Just like the proposed RL agent, the participants are allowed to hold only one unit of stock at a time. In the beginning, 30 days of stock historical prices are provided to the participant. Then, they are asked to begin trading with the objective of maximizing profit. A visual description of the task is shown in Fig. 5(*b*). They can *buy* and *sell* the stock like the proposed RL agent. Total profit and the profit made on the previous trade are displayed to the participants. The participants are asked to trade the bullish stock, the bearish stock, and the volatile stock used in Sub-section 2.3 for 180 days, in the same order.

After the completion of the simulated trading task for all the participants, the number of *average holding days* under three types of market conditions are shown in Table 4. In all market conditions, the number of average holding days of the risk-averse participants are much smaller than those of the risk-seeking participants. It is observed from Table 4 that risk-averse participants sell their stocks far more quickly than risk-seeking participants, in order to quickly cash out their profits.

Table 4. Number of average holding days of the participants under various market conditions

	Risk-averse participants	**Risk-seeking participants**
Bullish stock	12.42 days	46.66 days

Bearish stock	8.50 days	21.42 days
Volatile stock	11.16 days	16.67 days

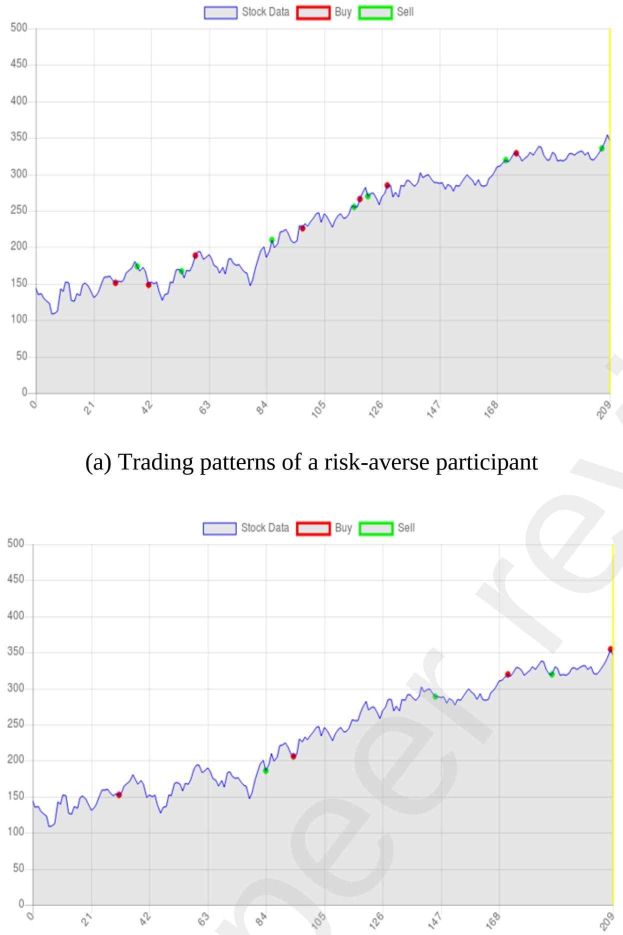
After the completion of the simulated trading task, the values of *percentage of holding period* are shown in Table 5 under various market conditions. The values of *percentage of holding period* of risk-averse participants are smaller than those of the risk-seeking participants. It can be observed from Table 5 that risk-averse participants are consistently less willing to hold their stocks compared to risk-seeking participants.

Table 5. *Percentage of holding period* of the participants under various market conditions

	Risk-averse participants	**Risk-seeking participants**
Bullish stock	48.36%	67.72%
Bearish stock	18.29%	36.53%
Volatile stock	22.89%	32.40%

These results in Table 4 and Table 5 match with the observations from Table 1 and Table 2, and support the analysis derived in Sub-section 2.3 with regards to the trading behaviors of the risk-averse RL agent and the risk-seeking RL agent. It has been validated that the trading pattern of the risk-averse participants is similar to that of the risk-averse RL agent. The risk-seeking participants perform stock trading with similar patterns of the risk-seeking RL agent.

The trading patterns for all 23 participants in three different market conditions were recorded. Two examples of trading patterns for the bullish stock by two participants are shown in Fig. 6, to better illustrate and compare the behavior differences of a risk-averse participant and a risk-seeking participant trading the same stock. It is compared against the observed from Fig. 6(*a*) that the risk-averse participant repeatedly opens and closes the positions in the bullish market and does not hold the positions for too long. It could be due to the fear of losing the profits. On the other hand, the risk-seeking participant does not sell too frequently and sticks to the position for longer durations, as shown in Fig. 6(*b*). It matches with the trading behavior observed in Sub-section 2.3 where the risk-averse RL agent consistently holds its stock for lesser number of days than those of the risk-seeking RL agent.



(a) Trading patterns of a risk-averse participant

(b) Trading patterns of a risk-seeking participant

Fig. 6. Examples of trading patterns for the same bullish stock.

Hence, according to experimental outcomes derived from the user study involving 23 participants using different risk taking tendencies; there is a behavioral similarity of the trading patterns between the participants and the proposed risk-awareness RL agent. It is a promising validation of the risk-awareness RL agent as it shows that the proposed RL agent does indeed behave like risk-averse and risk-seeking individuals in the real world. As such, the proposed risk-awareness RL agent has been validated and is ready to be incorporated within a risk-adaptive stock trading system. It will be described in the next section.

3. Proposed Risk-adaptive Stock Trading System

As observed from the annualized returns on investment by different risk-awareness agents under various market conditions in Table 3, different risk-sensitivity profiles are suitable under different stock market conditions. For example, when the market is bearish, being risk-averse behavior yields higher returns. When the market is bullish, exhibiting a risk-seeking behavior leads to higher returns. With the developed risk-awareness RL agent, a novel risk-adaptive stock trading system is proposed. The risk-adaptive stock trading system is capable of switching among appropriate risk-sensitivity profiles and judiciously select a suitable risk taking profile under the prevailing stock market conditions automatically.

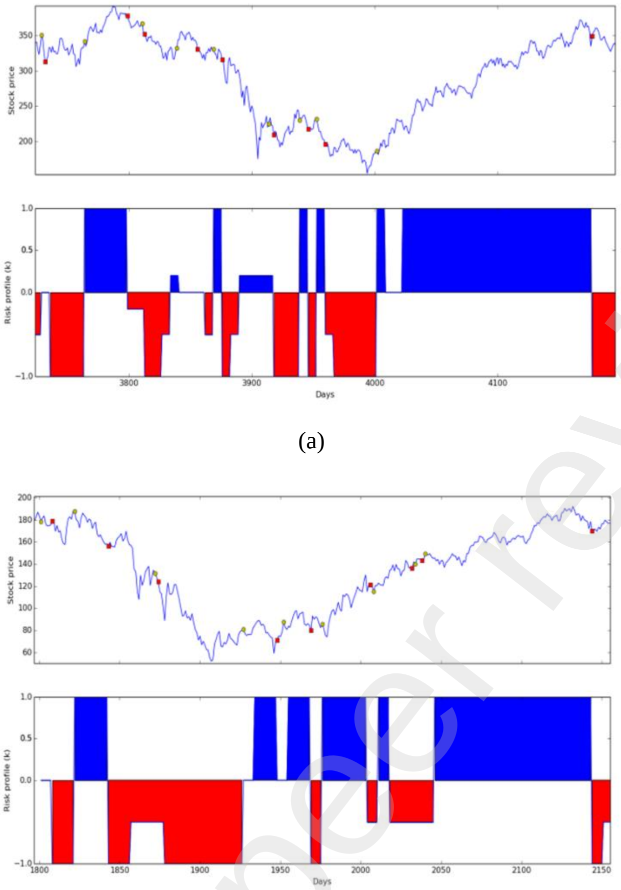
We need the trading system that switches to the risk-sensitivity to make a higher amount of profit. In order to do so, the developed trading system has a component called *AgentSelector*. It is configured by a parameter called *window_size* that is represented as the notation *ws*. There are seven pre-trained RL agents with different values of risk sensitivity parameter *k*, i.e., *k* = -1, *k* = -0.5, *k* = -0.2, *k* = 0, *k* = 0.2, *k* = 0.5, *k* = 1. The *AgentSelector* virtually performs trading with each agent for the past *ws* days, and selects the agent achieving the best profit in the past *ws* days. This scheme allows us to use the agent that has most recently been proven to be the best profitable. As the stock market conditions change, the agents with adifferent value of risk sensitivity *k* will become the most profitable. As such, under different market conditions, the *AgentSelector* will always identify the best performing agent and automatically switching to it for trading decision.

In order to evaluate the performance of the proposed risk-adaptive stock trading system, experiments are conducted on five volatile securities from the New York Stock Exchange (NYSE). It is to assess if the proposed trading system is capable of switching correctly risk-profiles at the suitable time. These five NYSE stocks are shown as follows.

- IShares MSCI Hong Kong Index Fund (EWH), in period of 16th April 2008 to 27th April 2010.
- Bank of America Corp (BAC), in period of 20th June 2008 to 15th Dec 2009.
- Goldman Sachs Group (GS), in period of 29th July 2008 to 8th January 2010.
- Citigroup Inc (C), in period of 11th July 2001 to 27th April 2004.
- JPMorgan Chase & Co. (JPM), in period of 11th July 2001 to 27th April 2004.

These five stocks and the time periods have been selected, such that each stock exhibits both bearish and bullish trends. The top half figures in Fig. 7 show the historical price charts of the examples of two stocks. It can be observed that the stocks' prices keep falling in the initial parts of the time series data. The stocks' prices are rising in the latter parts of the time series data. As a result, a fixed value of risk-sensitivity profile is not going to achieve the best trading profit. There is a need to automatically alter the risk profiles over time to gain more trading profit.

The trading patterns of the risk-adaptive stock trading system for these two example stocks are shown on the top half in Fig. 7. The bottom half in Fig. 8 shows the value of risk-sensitivity parameter *k* selected by the *AgentSelector* over time. The regions shaded in blue color show the periods where the risk-adaptive trading system chose a risk-seeking agent. The red regions show the periods where a risk-averse agent is chosen by the risk-adaptive trading system.



(a)

(b)

Fig. 7. Examples patterns of volatile stocks traded by the risk-adaptive trading system.

It is observed from Fig. 7 that the risk-adaptive trading system does indeed switch risk profiles over time in response to the underlying market conditions. It is evident that during bearish periods the risk-adaptive trading system uses risk-averse profiles. This is apparent from the fact the red colored shaded regions (i.e., representing risk-aversion) are significantly more prevalent in the initial parts of the price charts when stock prices are bearish. It is an optimal decision as during bearish times the risk-adaptive trading system avoids taking a position helps prevent losses due to falling stock prices. It is also evident that during bullish periods, the risk-adaptive system switches to risk-seeking profiles. This is detectable from the fact that blue colored shaded regions (i.e., representing risk-seeking) are dominant in the latter parts of the price charts when stock prices are bullish. This is an optimal decision because the system does not close its position too early and gains more profit when stock prices are rising.

Furthermore, the annualized returns on investment of the proposed risk-adaptive trading system are benchmarked against those of agents with immutable risk profiles and the MACD algorithm under volatile market conditions. The value of parameter *window_size* of the risk-adaptive system is set to 21. These five stocks are employed for trading by the proposed risk-adaptive trading system, a fixed risk-averse agent with risk sensitivity parameter $k = -0.5$, a fixed risk-neutral agent with parameter $k = 0$, a fixed risk-seeking agent with parameter $k = 0.5$, and the trading system using the MACD algorithm. The results of achieved returns are shown in Table 6. It is observed that the results of the risk-adaptive trading system are very encouraging, achieving 4.46% - 10.02% annualized returns on investment for the five stocks in the portfolio. The risk-adaptive trading system notably outperforms the RL agents with fixed risk-sensitivity profiles achieving -8.11% to 7.20% annualized returns on investment. It also outperforms the returns of the trading system on the MACD algorithm with -6.97% to 3.14% annualized returns on investment. It is because the risk-adaptive trading system is able to avoid losses during bearish periods by being risk-averse and earn better profits in bullish periods by switching to risk-seeking profiles automatically.

Table 6. Annualized returns on investment by risk-adaptive trading system compared to RL agents with fixed risk-sensitivities and the MACD algorithm

Stock name	**Risk-adaptive system**	**Fixed risk-averse agent (*k* = -0.5)**	**Fixed risk-neutral agent (*k* = 0)**	**Fixed risk-seeking agent (*k* = 0.5)**	**MACD algo.**
EWH	7.32%	2.92%	1.25%	-1.46%	0.88%
BAC	4.46%	-1.48%	-6.32%	-8.11%	-6.97%
GS	9.25%	3.18%	2.77%	3.13%	-3.85%
C	7.61%	4.47%	2.80%	5.33%	1.93%
JPM	10.02%	2.31%	5.79%	7.20%	3.14%

4. CONCLUSIONS

In this paper, our research works are presented in step-by-step manners. First, a RL model based on Q-learning is developed for stock trading. The RL agent is modelled and trained to identify troughs and peaks in time series stocks price charts. The agent is able to take actions to buy stock near troughs and off loading stock positions near peaks. The set of experimental results show that it maximizes the chances for profit making.

Next, the agent is further improved by incorporating risk factors to model different trading patterns undertaken by individuals with different risk-taking tendencies in the real world. The risk-awareness trading model is proposed to cover trading behaviors in the full spectrum of risk sensitivities and risk tendencies. It has a parameter k that allows for the setting of risk-sensitivity of the agent in an adjustable manner. In addition to being profit maximizing, it also takes into account the amount of risk the users are willing to tolerate. The experiments have been performed to evaluate its performance under four types of market conditions. The returns of investments achieved by the proposed risk-awareness agents with different risk profiles are compared to the trading model with the MACD algorithm.

Besides, the behavior of the risk-awareness trading model is successfully validated using the prospect theory. Another layer of affirmation is provided using the results of the user study tasks conducted by 23 participants with differing degree of risk appetites by comparing their trading patterns against those of the risk-awareness RL trading agent. The validation shows that the risk-awareness trading model indeed exhibits the anticipated behaviors for the different types of risk profiles. It successfully displays the desired trading behaviors. Both methods of validations are in agreement with the behaviors displayed by the risk-awareness trading agent.

Finally, this capability of the risk-awareness trading agent to manifest different risk profiles is leveraged to develop a risk-adaptive stock trading system that can adapt to the optimal risk strategy under different volatile stock market conditions. The risk-adaptive stock trading system is capable of taking optimal decisions, to automatically switch to be risk-averse during bearish times, and be risk-seeking during bullish periods. The results encouragingly show that the proposed risk-adaptive trading system achieves 4.46% - 10.02% annualized returns on investment for the five stocks in the portfolio. It outperforms the trading systems with immutable risk profiles from risk-averse, risk-neural, and risk-seeking. It also achieves higher returns than the trading model with the MACD algorithm.

ACKNOWLEDGMENT

The authors would like to thank Mr. Shantanu Gupta for his works, contributions, and help in this research.

REFERENCES

[1] W. Lu, J. Li, J. Wang, L. Qin, "A CNN-BiLSTM-AM method for stock price prediction," Neural Computing and Applications, vol. 33, pp. 4741–4753, 2021, https://doi.org/10.1007/s00521-020-05532-z.

[2] B. Alhnaity, M. Abbod, "A new hybrid financial time series prediction model," Engineering Applications of Artificial Intelligence, vol. 95, 2020, https://doi.org/10.1016/j.engappai.2020.103873.

[3] X. Li, C. Luo, "An intelligent stock trading decision support system based on rough cognitive reasoning," Expert Systems with Applications, vol. 160, 2020, https://doi.org/10.1016/j.eswa.2020.113763.

[4] Y. Kim, W. Ahn, K. Oh, D. Enke, "An intelligent hybrid trading system for discovering trading rules for the futures market using rough sets and genetic algorithms," Applied Soft Computing, vol. 55, pp. 127–140, 2017, https://doi.org/10.1016/j.asoc.2017.02.006.

[5] J. Ayala, M. García-Torres, J. Noguera, F. Gómez-Vela, F. Divina, "Technical analysis strategy optimization using a machine learning approach in stock market indices," Knowledge-Based Systems, vol. 225, 2021, https://doi.org/10.1016/j.knosys.2021.107119.

[6] E. Gerlein, M. McGinnity, A. Belatreche, S. Coleman, "Evaluating machine learning classification for financial trading: An empirical approach", Expert Systems with Applications, vol. 54, pp. 193-207, 2016. https://doi.org/10.1016/j.eswa.2016.01.018.

[7] R. Liem, "Momentum Investing Strategy in IDX: An Experiment," Journal of Applied Finance & Accounting, vol. 5, no. 1, pp. 71-109, 2012.

[8] B. Lim, S. Zohren, S. Roberts, "Enhancing Time-Series Momentum Strategies Using Deep Neural Networks," Journal of Financial Data Science, vol. 1, no. 4, pp. 19-38 2019, https://doi.org/10.3905/jfds.2019.1.015.

[9] Q. Huang, J. Yang, X. Feng, A. W. Liew, X. Li, "Automated Trading Point Forecasting Based on Bicluster Mining and Fuzzy Inference," IEEE Transactions on Fuzzy Systems, vol. 28, no. 2, 2020, https://doi.org/10.1109/TFUZZ.2019.2904920.

[10] Z. Tan, C. Quek, P. Cheng, "Stock trading with cycles: A financial application of ANFIS and reinforcement learning," Expert Systems with Applications, vol. 38, no. 5, 2011, https://doi.org/10.1016/j.eswa.2010.09.001.

[11] W.C. Chiang, D. Enke, T. Wu, R. Wang, "An adaptive stock index trading decision support system," Expert Systems with Applications, vol. 59, pp. 195–207, 2016, https://doi.org/10.1016/j.eswa.2016.04.025.

[12] D. Fengqian, L. Chao, "An Adaptive Financial Trading System using Deep Reinforcement Learning with Candlestick Decomposing Features," IEEE Access, vol. 8, 2020, https://doi.org/10.1109/ACCESS.2020.2982662.

[13] Y. Deng, F. Bao, Y. Kong, Z. Ren, Q. Dai, "Deep Direct Reinforcement Learning for Financial Signal Representation and Trading," IEEE Transactions on Neural Networks and Learning Systems, vol. 28, no. 3, pp. 653-664, 2017, https://doi.org/10.1109/TNNLS.2016.2522401.

[14] M. Alimoradi, A. Kashan, "A league championship algorithm equipped with network structure and backward Q-learning for extracting stock trading rules," Applied Soft Computing, vol. 68, 2018, https://doi.org/10.1016/j.asoc.2018.03.051.

[15] Y. Lin, S. Liu, H. Yang, H. Wu, B. Jiang B, "Improving stock trading decisions based on pattern recognition using machine learning technology," PLoS ONE, vol. 16, no. 8, 2021, https://doi.org/10.1371/journal.pone.0255558.

[16] L. Troiano, E. M. Villa and V. Loia, "Replicating a Trading Strategy by Means of LSTM for Financial Industry Applications," IEEE Transactions on Industrial Informatics, vol. 14, no. 7, pp. 3226-3234, July 2018, https://doi.org/10.1109/TII.2018.2811377.

[17] L. Conegundes, A. Pereira, "Beating the Stock Market with a Deep Reinforcement Learning Day Trading System," International Joint Conference on Neural Networks, pp. 1-8, 2020, https://doi.org/10.1109/IJCNN48605.2020.9206938.

[18] M. Aloud, N. Alkhamees, "Intelligent Algorithmic Trading Strategy using Reinforcement Learning and Directional Change," IEEE Access, 2021, https://doi.org/10.1109/ACCESS.2021.3105259.

[19] P. Pendharkar, P. Cusatis, "Trading financial indices with reinforcement learning agents," Expert Systems with Applications, vol. 103, pp. 1-13, 2018, https://doi.org/10.1016/j.eswa.2018.02.032.

[20] J. Chakole, M. Kolhe, G. Mahapurush, A. Yadav, M. Kurhekar, "A Q-learning agent for automated trading in equity stock markets," Expert Systems with Applications, vol. 163, 2021, https://doi.org/10.1016/j.eswa.2020.113761.

[21] Y. Yuan, W. Wen, J. Yang, "Using Data Augmentation Based Reinforcement Learning for Daily Stock Trading," Electronics, vol. 9, no. 9, 2020, https://doi.org/10.3390/electronics9091384.

[22] C. Kuo, C. Chen, S. Lin and S. Huang, "Improving Generalization in Reinforcement Learning–Based Trading by Using a Generative Adversarial Market Model," IEEE Access, vol. 9, 2021, https://doi.org/10.1109/ACCESS.2021.3068269.

[23] J. Carapuço, R. Neves, N. Horta, "Reinforcement learning applied to Forex trading," Applied Soft Computing, vol. 73, pp. 783-794, 2018, https://doi.org/10.1016/j.asoc.2018.09.017.

[24] C. Chen, A. Chen, S. Huang, "Cloning Strategies from Trading Records using Agent-based Reinforcement Learning Algorithm," IEEE International Conference on Agents, pp. 34-37, 2018, https://doi.org/10.1109/AGENTS.2018.8460078.

[25] E. Ponomarev, I. Oseledets, A. Cichocki, “Using Reinforcement Learning in the Algorithmic Trading Problem,” Journal of Communications Technology and Electronics, vol. 64, pp. 1450–1457, 2019, https://doi.org/10.1134/S1064226919120131.

[26] S. Almahdi, S. Yang, “A constrained portfolio trading system using particle swarm algorithm and recurrent reinforcement learning,” Expert Systems with Applications, vol. 130, pp. 145-156, 2019, https://doi.org/10.1016/j.eswa.2019.04.013.

[27] P. Gabrielsson and U. Johansson, "High-Frequency Equity Index Futures Trading Using Recurrent Reinforcement Learning with Candlesticks," IEEE Symposium Series on Computational Intelligence, pp. 734-741, 2015, https://doi.org/10.1109/SSCI.2015.111.

[28] J. Zhang, D. Maringer, "Using a Genetic Algorithm to Improve Recurrent Reinforcement Learning for Equity Trading", Computational Economics, vol. 47, pp. 551-567, 2016, https://doi.org/10.1007/s10614-015-9490-y.

[29] H. Park, M. Sim, D. Choi, “An intelligent financial portfolio trading strategy using deep Q-learning,” Expert Systems with Applications, vol. 158, 2020, https://doi.org/10.1016/j.eswa.2020.113573.

[30] Y. Li, P. Ni, V. Chang, “Application of deep reinforcement learning in stock trading strategies and stock forecasting,” Computing, vol. 102, pp. 1305–1322, 2020, https://doi.org/10.1007/s00607-019-00773-w.

[31] T. Théate, D. Ernst, “An application of deep reinforcement learning to algorithmic trading,” Expert Systems with Applications, vol. 173, 2021, https://doi.org/10.1016/j.eswa.2021.114632.

[32] B. Hirchoua, B. Ouhbi, B. Frikh, “Deep reinforcement learning based trading agents: Risk curiosity driven learning for financial rules-based policy,” Expert Systems with Applications, vol. 170, 2021, https://doi.org/10.1016/j.eswa.2020.114553.

[33] K. Lei, B. Zhang, Y. Li, M. Yang, Y. Shen, “Time-driven feature-aware jointly deep reinforcement learning for financial signal representation and algorithmic trading,” Expert Systems with Applications, vol. 140, 2020, https://doi.org/10.1016/j.eswa.2019.112872.

[34] G. Jeong, H. Kim, “Improving financial trading decisions using deep Q-learning: Predicting the number of shares, action strategies, and transfer learning,” Expert Systems with Applications, vol. 117, pp. 125-138, 2019, https://doi.org/10.1016/j.eswa.2018.09.036.

[35] H. Yang, X. Liu, S. Zhong, A. Walid, “Deep Reinforcement Learning for Automated Stock Trading: An Ensemble Strategy,” SSRN Electronic Journal, 2020, https://doi.org/10.2139/ssrn.3690996.

[36] K.E. Wärneryd, “Stock-Market Psychology: How People Value and Trade Stocks,” Publisher: Edward Elgar Publishing, 2001.

[37] P. Gai, N. Vause, “Measuring Investors' Risk Appetite,” Bank of England Working Paper Series No. 283, SSRN eJournal, 2005, http://dx.doi.org/10.2139/ssrn.872695.

[38] A. Díaz, C. Esparcia, “Assessing Risk Aversion from the Investor’s Point of View,” Frontiers in Psychology, vol. 10, 2019, https://doi.org/10.3389/fpsyg.2019.01490.

[39] O. Mihatsch, R. Neuneier, “Risk-Sensitive Reinforcement Learning,” Machine Learning, vol. 49, pp. 267–290, 2002, https://doi.org/10.1023/A:1017940631555.

[40] J. Li and L. Chan, "Reward Adjustment Reinforcement Learning for Risk-averse Asset Allocation," IEEE International Joint Conference on Neural Network, pp. 534-541, 2006, https://doi.org/10.1109/IJCNN.2006.246728.

[41] Y. Gao, Yue, Kry. Lui, P. Hernandez-Leal, “Robust Risk-Sensitive Reinforcement Learning Agents for Trading Markets,” poster session of International Conference on Machine Learning, 2021.

[42] T. Spooner, R. Savani, “Robust Market Making via Adversarial Reinforcement Learning,” International Joint Conference on Artificial Intelligence, Special Track on AI in FinTech, 2020.

[43] A. Tversky, D. Kahneman, “Advances in prospect theory: Cumulative representation of uncertainty,” Journal of Risk and uncertainty, vol. 5, no. 4, pp. 297-323, 1992.

[44] R.S. Sutton, A.G. Barto, “Reinforcement learning: An introduction,” Publisher: MIT press, 1998.

[45] V. Mnih, K. Kavukcuoglu, D. Silver, A. Rusu, J. Veness, et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. *518*, pp. 529-533, 2015, https://doi.org/10.1038/nature14236.

[46] T. Schaul, J. Quan, I. Antonoglou, D. Silver, “Prioritized experience replay,” arXiv preprint arXiv:1511.05952, 2015.

[47] H. Van Hasselt, A. Guez, D. Silver, “Deep Reinforcement Learning with Double Q-Learning,” 13th AAAI Conference on Artificial Intelligence, pp. 2094-2100, 2016.

[48] Y. Niv, J. Edlund, P. Dayan, J. O'Doherty, “Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain,” Journal of Neuroscience, vol. 32, no. 2, pp. 551-562, 2012, https://doi.org/10.1523/JNEUROSCI.5498-10.2012.

[49] J. O'Doherty, P. Dayan, K. Friston, H. Critchley, R. Dolan, “Temporal difference models and reward-related learning in the human brain,” Neuron, vol. 38, no. 2, pp. 329-337, 2003, https://doi.org/10.1016/s0896-6273(03)00169-7.

[50] S. Achelis, “Technical Analysis from A to Z,” Publisher: McGraw-Hill Education, 2nd edition, 2013.

[51] H. Boxer, “Moving Average Convergence/Divergence. Profitable Day and Swing Trading: Using Price/Volume Surges and Pattern Recognition to Catch Big Moves in the Stock Market,” Publisher: Wiley, 1st edition, 2014.

[52] P. Glimcher, E. Fehr, “Neuroeconomics: Decision making and the brain,” Publisher: Academic Press, 2nd edition, 2013.

[53] C. Lejuez, J. Read, C. Kahler, J. Richards, S. Ramsey, G. Stuart, D. Strong, R. Brown, “Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART),” Journal of Experimental Psychology Applied, vol. 8, no. 2, pp. 75, 2002, https://doi.org/10.1037//1076-898x.8.2.75.

[54] C. Lejuez, W. Aklin, S. Daughters, M. Zvolensky, C. Kahler, M. Gwadz, “Reliability and validity of the youth version of the Balloon Analogue Risk Task (BART–Y) in the assessment of risk-taking behavior among inner-city adolescents,” Journal of Clinical Child and Adolescent Psychology, vol. 36, no. 1, pp. 106-111, 2007, https://doi.org/10.1080/15374410709336573.

[55] D. van Ravenzwaaij, G. Dutilh, E. Wagenmakers, “Cognitive model decomposition of the BART: assessment and application,” Journal of Mathematical Psychology, vol. 55, no. 1, pp. 94-105, 2011, https://doi.org/10.1016/j.jmp.2010.08.010.