

Article

Analyzing Strategic Parental Leave Decisions Using Two-Player Multi-Agent Reinforcement Learning

Lixue Zhao  and Hyun-Rok Lee * 

Department of Industrial Engineering, Inha University, Incheon 22212, Republic of Korea; joyss77@inha.edu

* Correspondence: hyunrok.lee@inha.ac.kr; Tel.: +82-32-860-7373

Abstract

Despite the well-documented benefits of paid parental leave, many employees hesitate to take it. This study employs a two-player stochastic game (SG) model to analyze how various factors affect parental leave decisions. The proposed SG model incorporates (1) an employee's perceived utility from taking leave, (2) the effect of colleague's parental leave, (3) career penalties after taking leave, and (4) a paid parental policy. To accurately obtain equilibrium strategies, we extend Nash-Q learning by incorporating backward iteration and optimistic initialization. These two methods exploit the structural properties of the model to accelerate convergence and improve solution quality. Numerical experiments reveal that a stronger willingness to take parental leave and lower career penalties increase parental leave uptake. Furthermore, the competitive career penalty, which captures interpersonal factors, is particularly influential when a colleague is less likely to take parental leave. Our results suggest that reducing career penalties can substantially increase leave uptake in typical parameter ranges, highlighting the importance of workplace policies that mitigate career penalties associated with parental leave.

Keywords: paid parental leave; stochastic game; career penalty; Nash Q-learning; multi-agent reinforcement learning

1. Introduction

In recent years, paid parental leave (PPL) policies have been recognized as an important means of enabling employees to strengthen family relationships and achieve better work–life balance. By providing financial support, PPL policies enable employees to actively engage in childcare while maintaining their professional careers [1–3]. They can also contribute to the physical and mental health of both parents and children [4–6]. Moreover, PPL policies can positively affect the labor market [7,8] and gender equality [9,10]. However, despite these numerous benefits, many employees hesitate to take parental leave, even when they genuinely want to.

Previous studies show that decision factors for taking parental leave include environmental factors such as social norms [11–13], company culture [14], and family support [15], as well as individual factors [16–18]. For instance, Carlsson [11] empirically finds that coworkers' usage of parental leave can positively influence the usage of parental leave through an instrumental variable approach called 'peers of peers'. Similarly, other studies have also found that individuals may imitate the behavior of peers to conform to social norms within their social group or workplace [12,13]. On the other hand, survey-based evidence suggests that parental leave can negatively affect employees' job commitment, and that this negative perception can be alleviated with parental-leave-friendly organizational



Academic Editor: Fernando De la Prieta Pintado

Received: 16 January 2026

Revised: 13 February 2026

Accepted: 14 February 2026

Published: 19 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

policies [14]. Hence, to analyze the rationale behind parental leave decisions, an ecological model that captures multi-level influences—including interpersonal, organizational, community, and public policy levels—is necessary, as has been similarly used in analyzing health behavior [19].

This study examines multi-level influences on parental leave decisions through a game-theoretic framework. We analyze how such decisions differ in competitive work environments. Specifically, we formulate a stochastic game (SG) model to represent competitive dynamics between two employees who begin their work careers with the same conditions. Each employee in the model optimizes whether to take parental leave each year considering (1) the personal valuation of parental leave (individual factor), (2) the parental leave usage of the other employee (interpersonal factor), (3) penalties for promotion after taking parental leave (organizational and community factors) and (4) financial support during leave (public policy factor). Accordingly, solutions to the proposed SG model characterize the equilibrium behaviors of the two employees under different multi-level influences. Thus, the formulated SG model is used to answer the following key research questions:

- In a competitive environment, does the presence of an additional career penalty mechanism discourage employees from taking parental leave?
- How does the magnitude of the career penalty affect the employees' optimal strategies?
- Is an employee's leave decision influenced by the other employee's decision? For example, if one employee does not take parental leave, is the other more likely to forgo leave to maintain a competitive advantage?

Existing work has applied reinforcement learning to two-player game-theoretic scenarios [20–22], demonstrating the effectiveness of Q-learning approaches in multi-agent settings. To obtain and identify equilibrium behaviors of two employees, we propose a variant of the Nash Q-learning algorithm [20] that is integrated with the backward iteration method. This solution method leverages the structural properties of the SG model to accelerate convergence to the optimal solution, ensuring that the derived solutions correctly capture expected behaviors of employees. Through the formulated model and the proposed solution method, this study provides insights into how organizational policies and cultural norms shape employees' parental leave decisions. The contribution of this study is threefold:

- It introduces a game-theoretic framework that models the strategic decision-making process of two employees regarding parental leave, accounting for both individual career goals and inter-agent competition;
- It identifies optimal strategies for employees under varying conditions of income replacement and career penalties using an SG model and the Nash Q-learning algorithm;
- It explores how changes in income replacement and career penalties affect parental leave decisions, with implications for organizational policy design.

2. Literature Review

Parental leave has been widely studied in different disciplines, including sociology, economics, and organizational studies. Existing studies focus mainly on two aspects: the decision factors for parental leave usage and its consequences. To systematically review the decision factors affecting parental leave usage, we employ the ecological model, which emphasizes multi-level influences on behavior [19]. The ecological model categorizes the influences into five levels: individual, interpersonal, organizational, community, and public policy. At the individual level, personal characteristics such as attitudes toward parental leave and income replacement have been identified as important influences on parental leave behavior. Romero-Balsas et al. [16], through in-depth interviews with 30 Spanish fathers and critical discourse analysis, found that fathers who took longer

parental leave view taking parental leave as a responsibility toward their families as well as a personal right. Willingness to engage in housework and childcare or to pursue work–family balance has also been recognized as important factors in parental leave decisions in South Korea [17,18]. Additionally, income replacement has been frequently highlighted as a major decision factor in parental leave decisions. Doucet and McKay [9] identified that income and the length of leave were important decision factors based on interviews with 26 couples in Canada. Kaufman [23] conducted interviews in the UK and found that insufficient income replacement often discourages fathers from taking parental leave. Similarly, Sogaard and Jørgensen [24] found that higher income replacement significantly increases the likelihood of fathers taking parental leave using Danish registry data covering approximately 200,000 birth cases.

At the interpersonal level, family members and coworkers can affect parental leave decisions. For example, McKay and Doucet [15] conducted interviews with 26 couples and found that fathers' parental leave decisions were influenced by mothers' preference, which shaped both the timing and duration of fathers' leave. Carlsson [11] found that an increase in the parental leave duration taken by coworkers is associated with a corresponding increase in the parental leave duration taken by other employees. On average, a 10-day increase in the duration of parental leave taken by coworkers led to an increase of approximately 1.5 days for male colleagues and 1 day for female colleagues. At the organizational level, previous studies show that company culture significantly influences the usage of parental leave. Employees often report workplace pressure and career concerns as barriers to taking leave, even when legally entitled [13,25–27].

At the community and public policy level, existing parental leave policies affect parental leave decisions. O'Brien [2], through an analysis of parental leave policies in 24 countries, found that fathers' parental leave usage increases when policies are designed for fathers and ensure higher income replacement. Margolis et al. [28] analyzed the changes in parental leave usage rates following two parental leave policies in Canada in 2006 and found that different parental leave policies can affect parental leave behavior depending on income replacement levels. Thus, numerous studies have explored the decision factors associated with parental leave across different levels of influence. However, most of these studies are empirical and retrospective, as they have primarily relied on interview data and historical records. In contrast, this study employs a mathematical model and simulation experiments to predict parental leave-taking behaviors shaped by individual, interpersonal, organizational, community, and policy-level influences.

The second line of parental leave research has examined the effects of parental leave on professional careers, mental health, gender equality, and other outcomes. Many studies have identified negative effects of parental leave on employees' professional careers [14,29–32]. Absences due to parental leave can significantly affect employees' careers, especially in terms of pay and promotions [29,30]. Evertsson [31] found a negative impact of parental leave on both men's and women's wages using Swedish population registration data from 1990 to 2009. In particular, women's wage growth significantly declines after taking parental leave, as it often leads to missed promotion opportunities and associated salary increments. These negative outcomes of parental leave can be interpreted through signaling theory, which implies that taking parental leave may create negative perceptions of an employee's job commitment, thereby leading to fewer opportunities for wage increases or promotions [32]. Petts et al. [14], based on a survey, found empirical evidence of such negative perceptions. Research on the impact of parental leave on family well-being has primarily focused on maternal mental health [4,33,34] and child development outcomes [6,35–38]. These studies have reported that parental leave lowers stress levels and increases job satisfaction among mothers, while children also tend to exhibit better

cognitive and emotional outcomes. Other studies have identified positive effects of fathers' parental leave on overall family well-being [2,39,40]. Moreover, parental leave policies can contribute to gender equality by encouraging fathers to take leave—particularly through policies such as “use-it-or-lose-it” quotas and “daddy-month”—which have been shown to reduce gender disparities in career interruptions [41–43]. Building on these findings, this study formulates an SG model to explore how organizations can foster parental-leave-friendly environments that effectively leverage the well-documented benefits of parental leave.

3. Mathematical Model

3.1. Preliminary: Stochastic Game (SG) Model

A stochastic game (SG) [44] is a multi-agent extension of the Markov decision process (MDP), in which multiple agents simultaneously interact in a shared environment. A two-player stochastic game is defined as a 6-tuple $\langle S, A, r, p, \gamma, T \rangle$. S denotes the state space and $A = A^1 \times A^2$ denotes the agents' joint action space. At each time step, agents observe the current state $s_t \in S$ and choose their actions $a_t = (a_t^1, a_t^2) \in A(s_t)$. The individual policy function $\pi^i(a_t|s_t) \in [0, 1]$ for $i \in \{1, 2\}$ defines the probability of selecting a_t when an agent i observes state s_t . The joint action $a_t = (a_t^1, a_t^2)$ determines the next state s_{t+1} according to the transition probability $p(s_{t+1}|s_t, (a_t^1, a_t^2))$ and each agent receives an individual reward $r^i(s_t, a_t)$, for $i \in \{1, 2\}$. This process repeats until the time horizon T is reached. In a discounted stochastic game, the objective of each player is to maximize the discounted sum of rewards, with discount factor $\gamma \in [0, 1]$. The solution to an SG model is a Nash equilibrium policy $\pi^* = (\pi_*^1, \pi_*^2)$, where each agent's strategy is a best response to the other agent's strategy. The value functions under policy π^* for agent $i \in \{1, 2\}$ are defined as follows [20]:

$$V_*^i(s) = V_{(\pi_*^1, \pi_*^2)}^i = \mathbb{E}_{s_t, (a_t^1, a_t^2) \sim \pi^*} \left[\sum_{t=1}^T \gamma^t r^i(s_t, (a_t^1, a_t^2)) \mid s_1 = s \right] \quad (1)$$

$$Q_*^i(s_t, (a_t^1, a_t^2)) = r^i(s_t, (a_t^1, a_t^2)) + \gamma \cdot \sum_{s_{t+1} \in S} p(s_{t+1}|s_t, (a_t^1, a_t^2)) V_*^i(s_{t+1}) \quad (2)$$

A Nash equilibrium policy π^* satisfies the following conditions, where π_*^{-1} (π_*^{-2}) denotes the policy of agent 2 (agent 1) deviating from the π^* .

$$\begin{aligned} V_*^1(s) &\geq V_{(\pi_*^1, \pi_*^{-1})}^1(s) \quad \forall \pi_*^{-1} \\ V_*^2(s) &\geq V_{(\pi_*^{-2}, \pi_*^1)}^2(s) \quad \forall \pi_*^{-2} \end{aligned} \quad (3)$$

It means that no agent has an incentive to deviate from the Nash equilibrium policy given the strategies of the other agents.

3.2. Problem Description

This study considers a target problem in which two employees begin their careers in the same job position, which represents a competitive work environment. Each employee aims to maximize the returns over the total work period of T years. They decide whether to take parental leave each year, considering their income and the possibility of future promotions. For tractability, we consider a simplified setting with one child and a single one-year parental leave decision. The annual salary $f(x) \in \mathbb{R}^+$ depends on the job position x , and $f(x)$ increases with higher positions. There exist $\bar{x}$ job positions, and a larger x value indicates a higher position. At each job position x , an employee becomes eligible for promotion after completing the minimum required number of years $y_{p,x}$. Thereafter, the

employee receives one promotion opportunity per year. If an employee continues to miss promotion opportunities and reaches $\bar{y}_{p,x}$ years in the current position, the employee will no longer be eligible for promotion and will remain in the current position until retirement. The probability of promotion $q(x, m^i, m^{-i})$ depends on the current position x , whether the employee has taken a parental leave before (indicated by m^i), and whether the other employee has taken a parental leave before (indicated by m^{-i}). This probability decreases with higher-level positions:

$$q(x, m^i, m^{-i}) \leq q(y, m^i, m^{-i}) \quad \text{if } x \geq y.$$

There exist two types of career penalties that reduce the promotion probability. The first type of penalty $\delta \geq 0$ represents the individual career penalty for taking parental leave, defined as the reduction in promotion probability after taking parental leave.

$$\delta = q(x, 1, 0) - q(x, 0, 0)$$

The second type of penalty $\alpha \geq 0$ is the competitive career penalty, which is activated only when the employee has taken parental leave while the other employee in the same job position has not and competes for a promotion.

$$\alpha = q(x, 0, 0) - q(x, 0, 1)$$

According to this specification, the ordering of promotion probabilities is $q(x, 1, 1) = q(x, 1, 0) > q(x, 0, 0) > q(x, 0, 1)$. These probabilities are given by $q(x, 1, 1) = q(x, 1, 0) = q(x)$, $q(x, 0, 0) = q(x) - \delta$, and $q(x, 0, 1) = q(x) - \delta - \alpha$. For example, if $q(x) = 0.65$, $\delta = 0.10$, $\alpha = 0.05$, then $q(x, 1, 1) = q(x, 1, 0) = 0.65$, $q(x, 0, 0) = 0.55$, $q(x, 0, 1) = 0.50$. The individual career penalty for taking parental leave can discourage employees from using it unless sufficient income replacement is guaranteed. Moreover, the competitive career penalty creates a strategic dilemma in which both employees may choose to avoid parental leave in order to remain competitive in the workplace. As previous studies have found that career penalties associated with parental leave arise from negative perceptions, when an employee has not taken parental leave, the promotion probability is assumed to remain unchanged regardless of the colleague's parental leave usage. Through simulation experiments, this study examines how the magnitude of these penalties influences parental leave-taking behavior. The eligibility criteria for parental leave and the subsidy structure in this study are based on South Korean labor policies. Employees are allowed to take up to one year of paid parental leave per child before the child reaches the age of eight. During the leave period, the government provides financial support that depends on the employee's annual salary. The subsidy amount is subject to both upper and lower bounds. The leave period is counted toward years of service, as guaranteed by law. However, it is assumed that employees are not eligible for promotion while they are on parental leave.

3.3. Formulation

The target problem is formulated as a two-player SG model $\langle S, A, r, p, \gamma, T \rangle$. The model captures the transition dynamics and reward outcomes when employees make parental leave decisions each year.

- State $s_t = (s_t^1, s_t^2) \in S$. The state at year $t \in \{1, \dots, T\}$ consists of each employee's individual state $s_t^i = (x^i, y_p^i, m^i, y_c^i)$. Here, $x^i \in \{1, 2, \dots, \bar{x}\}$ is the employee's current job position; $y_p^i \in \{0, 1, \dots, y_{p,x^i}, \dots, \bar{y}_{p,x^i}\}$ is the number of years spent in the current position; $m^i \in \{0, 1\}$ is an indicator of whether the employee is eligible for parental leave, where $m^i = 1$ means that the employee has not yet used parental leave and

remains eligible, and $m^i = 0$ means that the employee has already used parental leave and is no longer eligible; and $y_c^i \in \{0, 1, \dots, \bar{y}_c\}$ is the child's age. If $y_c^i = \bar{y}_c$, the employee is no longer eligible to take parental leave, even if it has not been used previously.

- Joint action $a_t = (a_t^1, a_t^2) \in A(s_t)$. Employees can choose either to continue working (W) or to take one year of parental leave (L); $a_t^i \in \{W, L\}$, if they are eligible to take parental leave. If an employee has already used parental leave ($m^i = 0$) or the child has reached the maximum eligible age ($y_c^i = \bar{y}_c^i$), then the employee must work; i.e., $a_t^i \in \{W\}$.
- Individual reward function $r^i(s_t, a_t) \in \mathbb{R} \forall s_t \in S, a_t \in A(s_t)$. The reward received by an employee in each year depends on the employee's current position x^i and the chosen action a_t^i .

$$r^i(s_t, a_t) = f(x^i) \cdot 1_{a_t^i=W} + \{g(f(x^i)) + U_i^+\} \cdot 1_{a_t^i=L} \quad (4)$$

where $g(f(x)) \in \mathbb{R}^+$ denotes the government subsidy provided to an employee when they take parental leave, based on the annual salary $f(x)$, and $U_i^+ \in \mathbb{R}^+$ represents the employee's perceived utility from taking one year of parental leave. Although the perceived utility value captures non-monetary benefits of taking parental leave, we assume that U_i^+ is defined on a cardinal scale that is commensurate with income, thereby allowing algebraic aggregation. Accordingly, this value should be interpreted in monetary-equivalent terms rather than as a direct psychological measure.

- Transition probability $p(s_{t+1}|s_t, a_t) \in [0, 1] \forall s_t \in S, a_t \in A(s_t), s_{t+1} \in S$. If $a_t^i = L$, the next individual state s_{t+1}^i is determined with probability one. The job position x^i remains unchanged; y_p^i increases by 1, up to the maximum $\bar{y}_{p,x^i}$; m^i becomes 0; and y_c^i increases by 1, up to $\bar{y}_c$. Otherwise, if $a_t^i = W$, the employee may be promoted to the next position $x^i + 1$ with promotion probability $q(x^i, m^i, m^{-i})$ provided that the years of service at the current position meet the promotion eligibility condition, i.e., $(y_{p,x^i} \leq y_p^i < \bar{y}_{p,x^i})$. If the employee is not promoted or is outside the promotion-eligible years, the state transition follows the same logic in the case of taking parental leave, except that m^i remains unchanged.
- Discount factor $\gamma \in [0, 1]$. In this model, γ reflects an annual interest rate.
- Time horizon $T \in \mathbb{Z}^+$. T denotes the maximum number of service years of an employee.

In summary, the proposed model integrates influential factors on parental leave decisions across multiple levels. The individual level factor is captured by the perceived utility U^+ value, which represents an employee's intrinsic valuation of taking parental leave. Interpersonal and organizational influences are modeled through the competitive career penalty α and the individual career penalty δ , respectively. Finally, the subsidy function $g(f(x))$ represents the public policy level factor.

4. Algorithm

4.1. Preliminary: Nash Q-Learning

Nash Q-learning can find a solution to a stochastic game that satisfies the conditions of Equation (1) [20]. It is a sample-based algorithm that repeatedly obtains a transition sample (s_t, a_t, r_t, s_{t+1}) according to the SG model and updates the Q-values $Q^i(s_t, (a_t^1, a_t^2))$ for agent $i \in \{1, 2\}$ as follows:

$$Q^i(s_t, a_t) = (1 - l_n)Q^i(s_t, a_t) + l_n[r^i(s_t, a_t) + \gamma \text{Nash}Q^i(s_{t+1})] \quad (5)$$

where l_n is the learning rate at the n^{th} update. At the current state s_t , each agent i selects its action a_t^i according to an ϵ -greedy policy, which chooses a Nash equilibrium action $a_t^{i,NE}$ with probability $1 - \epsilon + \epsilon/|A^i|$:

$$a_t^i = \begin{cases} a_t^{i,NE}, & \text{with probability } 1 - \epsilon + \epsilon/|A^i| \\ \text{random action} & \text{with probability } \epsilon/|A^i| \end{cases}$$

where $|A^i|$ is the number of actions available to agent i . A Nash equilibrium at state s_t is identified by constructing a stage game based on the Q-values of all possible joint actions. These values are converted into payoff matrices for each agent and used to define the stage game. The Nash equilibrium strategies (π_t^1, π_t^2) are then computed using the Lemke-Howson algorithm [45]. This algorithm selects one Nash equilibrium from the possibly multiple Nash equilibria of the stage game, and in our implementation, one equilibrium is randomly selected. Finally, $NashQ^i(s_t)$ is defined as the expected Q-value of agent i under the selected Nash equilibrium strategies:

$$NashQ^i(s_t) = \sum_{a^1, a^2} \pi_t^1(a^1) \cdot \pi_t^2(a^2) \cdot Q^i(s_t, a^1, a^2).$$

Thus, Nash Q-learning solves the stage game at every transition and updates the Q-values based on the Nash equilibrium of the game. The algorithm is theoretically guaranteed to converge to the optimal Q-values as defined in Equation (2). The pseudo-code of the Nash Q-learning algorithm is provided in Appendix A.3.

4.2. Accelerating Convergence via Optimistic Initialization and Backward Iteration

Despite its theoretical guarantee of convergence, vanilla Nash Q-learning often requires an excessive number of transition samples to converge and sometimes inaccurately estimates the optimal Q-values for the formulated SG model. These issues may be due to the model's large state space and the limitations of sample-based Q-value approximations. Therefore, to improve the convergence speed of Nash Q-learning, this study introduces two methods: optimistic initialization and backward iteration.

The first method, optimistic initialization, initializes the Q-values using the optimal Q-values obtained from a single-agent model that does not have the competitive career penalty. The single-agent model is a Markov decision process (MDP) model defined by an individual state, individual action, individual reward, and the corresponding transition dynamics. Note that the state of the proposed SG model is a combination of the individual states of each agent, i.e., $s_t = (s_t^1, s_t^2)$. Transitions of an individual state are independent of the other agent's state except for the influence of the competitive career penalty. Therefore, in the absence of this penalty, we can define separate MDP models for agent 1 and agent 2. A detailed description of the single-agent MDP model is provided in Appendix A.1. The optimal Q-values for these single-agent models are likely to be larger than those of the two-agent model, as one source of career penalty is removed. Optimistic initialization leverages this property to encourage exploration of unseen state-action pairs by setting the initial Q-values using the optimal Q-values of the single-agent models. In Nash Q-learning integrated with optimistic exploration, unexplored state-action pairs are more likely to be selected because they initially have overestimated Q-values, which are later corrected through actual experience. Furthermore, dynamic programming methods such as value iteration and policy iteration can obtain exact optimal Q values for single-agent models. These exact values coincide with the optimal Q-values of the two-agent model for certain states and also help to accurately estimate Q-values of other states as well.

The second method, backward iteration, defines the learning order over states of interest. The states of interest are those in which both employees are in the same status and are eligible to take parental leave; that is, $s_t = (s_t^1, s_t^2)$ such that $s_t^1 = s_t^2 = (x, y_p, m = 1, y_c)$, where $x \in \{1, 2, \dots, \bar{x} - 1\}$, $y_p \in \{0, 1, \dots, y_{p,x}, \dots, \bar{y}_{p,x}\}$, and $y_c \in \{0, 1, \dots, \bar{y}_c - 1\}$. This set is defined to examine how the parental leave behaviors of two employees vary under different work conditions when they begin their careers in the same status. Backward iteration exploits the structural characteristics of the proposed SG model to determine which state in the set of states of interest should be examined first. In this model, for any given state, its possible predecessor and successor states are mutually exclusive. If one state can be reached after another, the reverse visitation is not possible, as the values of each state component either monotonically increase or decrease. For example, an employee's current job position x^i cannot decrease during transitions. Nash Q-learning with backward iteration begins learning from the state that can be visited last among the states of interest. After completing a sufficient number of episodes, the algorithm proceeds to the next state in the reverse visitation order and continues learning. Although this approach still relies on sample-based approximation, Q-value estimation becomes more efficient and accurate by utilizing previously learned Q-values of successor states. Details of Nash Q-learning with optimistic initialization and backward iteration are provided in Algorithm 1.

Algorithm 1 Nash Q-Learning with optimistic initialization and backward iteration.

```

1: Initialize:  $Q^1, Q^2$ 
2: Initialize the list of the states of interest  $S_0$  and the list of the studied states  $S_{\text{done}}$ 
3: for each  $s_0 \in \text{reversed}(S_0)$  do
4:   Initialize  $n \leftarrow 0$ 
5:   Initialize  $\epsilon_n, l_n$  ▷ Exploration and learning rate
6:   for the total number of episodes do
7:      $s_t \leftarrow s_0$ , episode  $\leftarrow []$ 
8:     while episode not terminated do
9:       for  $i \in \{1, 2\}$  do
10:        if  $Q^i((s_t^1, s_t^2), (a_t^1, a_t^2))$  undefined then
11:           $Q^i((s_t^1, s_t^2), (a_t^1, a_t^2)) \leftarrow Q_{\text{single}}(s_t^i, a_t^i)$ 
12:        end if
13:      end for
14:      if  $s_t \notin S_{\text{done}}$  then
15:        Select  $(a_t^1, a_t^2)$  via  $\epsilon$ -greedy policy based on  $Q^1, Q^2$ 
16:        Execute actions, observe  $(r_t^1, r_t^2), s_{t+1}$ 
17:        Append  $(s_t, (a_t^1, a_t^2), (r_t^1, r_t^2), s_{t+1})$  to episode
18:      else
19:        break
20:      end if
21:       $s_t \leftarrow s_{t+1}$ 
22:    end while
23:    for each  $(s_t, (a_t^1, a_t^2), (r_t^1, r_t^2), s_{t+1})$  in reversed(episode) do
24:      Update  $Q^1, Q^2$  by Equation (5)
25:    end for
26:     $n \leftarrow n + 1$ 
27:    if  $n \bmod 10 = 0$  then
28:      update  $\epsilon_n, l_n$ 
29:    end if
30:  end for
31:  Append  $s_0$  to  $S_{\text{done}}$ 
32: end for

```

5. Numerical Experiment

5.1. Experimental Settings

Numerical experiments are designed to demonstrate a typical work environment in South Korea. Employees are assumed to work for $T = 25$ years. There are $\bar{x} = 5$ job positions, in ascending order: staff ($x = 1$), assistant manager ($x = 2$), manager ($x = 3$), senior manager ($x = 4$), and director ($x = 5$). For each job position, the annual salary $f(x)$,

the minimum service years for promotion eligibility $y_{p,x}$, and the promotion probability without a career penalty $q(x, 1, \cdot)$ are inferred from data provided by the Wage and Job Information System of the Ministry of Employment and Labor of Korea and the wage system and the workforce management survey report of 2023 [46,47]. The number of promotion opportunities is set to 2 at each job position; hence, $\bar{y}_{p,x} = y_{p,x} + 3$. The assumed values regarding salary structure are summarized in Table 1. Employees are eligible to take one year of paid parental leave if their child is aged 8 or younger; that is, $y_c \in \{0, 1, \dots, 9\}$, where $y_c = 9$ indicates that the child has exceeded the eligibility age. Under the current paid parental leave policy in South Korea (as of 2025), the government subsidy is set at 80% of the employee's annual salary, subject to a minimum of 8.4 million KRW and a maximum of 21.5 million KRW per year. Accordingly, $g(f(x)) = \max\{8.4 \text{ M}, \min\{0.8f(x), 21.5 \text{ M}\}\}$. The discount factor is set to 0.95, assuming an annual interest of 5%.

Table 1. Salary structure assumed in numerical experiments.

Position x	Staff ($x = 1$)	Assistant Manager ($x = 2$)	Manager ($x = 3$)	Senior Manager ($x = 4$)	Director ($x = 5$)
Annual salary $f(x)$ (KRW)	30 M	36 M	47 M	53 M	66 M
Minimum service years $y_{p,x}$	3	4	4	5	5
Base promotion probability $q(x, 1, \cdot)$	1.0	0.65	0.45	0.35	–

Numerical experiments consist of two settings. (1) The first setting examines scenarios in which both employees have the same perceived utility from taking parental leave ($U_1^+ = U_2^+$). (2) The second setting considers scenarios in which employees perceive different values for parental leave ($U_1^+ \neq U_2^+$). In both settings, simulated experiments are conducted across combinations of $\delta \in \{0, 0.1\}$ and $\alpha \in \{0, 0.05, 0.1\}$. Note that $\delta = 0$ indicates that there is no individual career penalty for taking parental leave, and $\alpha = 0$ indicates the absence of a competitive career penalty. Thus, these settings allow us to examine how employees' strategic behavior changes depending on the presence or absence of each penalty. For the same-utility settings, utility values are selected from the set $\{0, 33, 41.5, 50, 100, 200\}$ (millions). In addition, a setting with $(\alpha = 0, \delta = 0.2)$ is included to further investigate the effect of δ when no competitive penalty exists. For the different-utility settings, combinations of utility values are chosen from the set $\{(0, 33), (0, 50), (0, 100), (33, 50), (33, 100), (100, 50)\}$ (millions). Results from these settings, along with the same-utility settings, allow us to compare how employees' strategic behaviors change when U_2^+ varies from 0, 33, 50 to 100 million for a fixed $U_1^+ \in \{0, 33, 50\}$. This includes scenarios where an employee does not take parental leave at all ($U_1^+ = 0$), as well as scenarios in which one employee's perceived utility remains fixed while the other's is relatively lower or higher.

In our implementation of the Nash-Q learning algorithm, the exploration rate (ϵ_0) is set to 0.99 and the initial learning rate (l_0) is set to 0.5. Both parameters decay at a rate of 0.99 every 10 episodes to balance exploration and convergence. The minimum values are 0.50 for ϵ and 0.05 for the learning rate. Each state of interest is trained for 5000 episodes. As an evaluation metric, we count how frequently each agent takes parental leave before retirement when both agents begin their careers from the same state. For each state of interest, we generate 100 episodes and record whether each agent takes parental leave before termination. The same random seed is used across different parameter settings and utility values to ensure a fair comparison. We then aggregate the number of parental leave cases across all states of interest and divide it by the total number of episodes to define the parental leave probability. This probability of an agent, which serves as the main metric in the experimental results, represents the proportion of episodes in which the agent takes parental leave before retirement. Experiments were run on an Intel Core i9-13900K CPU

(Intel Corporation, Santa Clara, CA, USA) with 128 GB RAM, and the maximum runtime was 31 h. The full implementation of our method is provided in Appendix B.

5.2. Results and Discussion

5.2.1. Results Under Equal Utility Values

Figure 1 shows how the parental leave probability of an employee changes with different utility levels, under the condition that both employees have the same perceived utility values and no competitive penalty ($\alpha = 0$). We plot the parental leave probabilities of employee 2 without loss of generality, as employee 1 exhibits similar behavior under the equal-utility level settings. The full results of the experiments are provided in Appendix A.2. The parental leave probability of an employee increases as the perceived utility value increases (reflecting the influence of individual factors). In particular, when the utility value reaches 200 M, the employee almost always takes parental leave, even as δ increases. This suggests that when perceived utility is sufficiently high, employees are willing to take parental leave despite potential penalties in promotion opportunities. This finding is consistent with previous studies that emphasize the role of individual factors such as the prioritization of family relationships and the adequacy of income replacement. For the same utility values, the parental leave probability decreases as δ increases (reflecting the influence of organizational factors). Employees become more reluctant to take parental leave because higher promotion penalties lead to greater losses in expected career returns after taking parental leave. This behavior is more evident at moderate utility levels, where whether the income replacement exceeds the income replacement threshold for taking parental leave is highly sensitive to environmental factors.

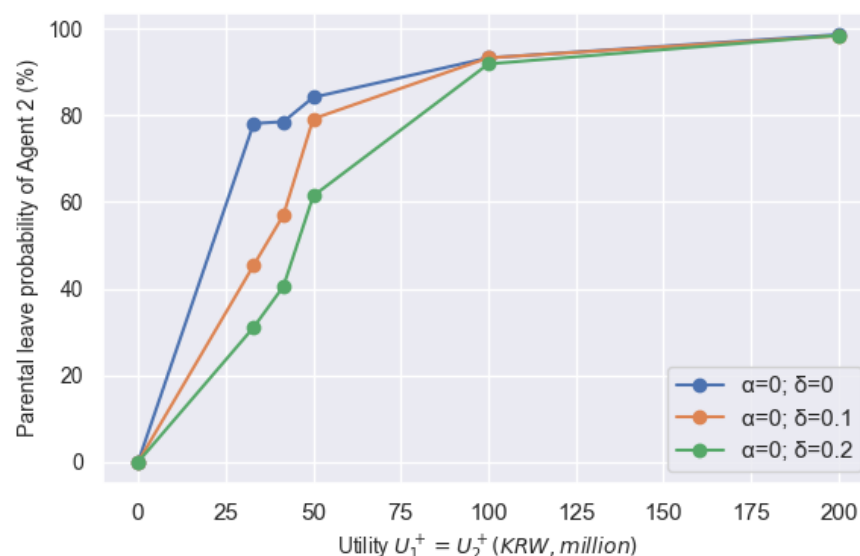


Figure 1. Comparison of the parental leave probability of agent 2 when $U_1^+ = U_2^+$, $\alpha = 0$ and δ varies among $\{0, 0.1, 0.2\}$. Higher δ significantly reduces leave uptake when U_2^+ is moderate but has negligible effect when U_i^+ is very high.

On the other hand, the impact of the competitive penalty (α) on the probability of taking parental leave is minimal when both employees have the same utility values (Figure 2). For both levels of δ , different values of α do not lead to significant changes in parental leave behavior. This is because both employees tend to take parental leave together or neither employee takes leave when they begin their careers from the same state and have the same perceived utility. Hence, the effect of α is highly limited, as situations in which one employee takes leave while the other does not occur only rarely.

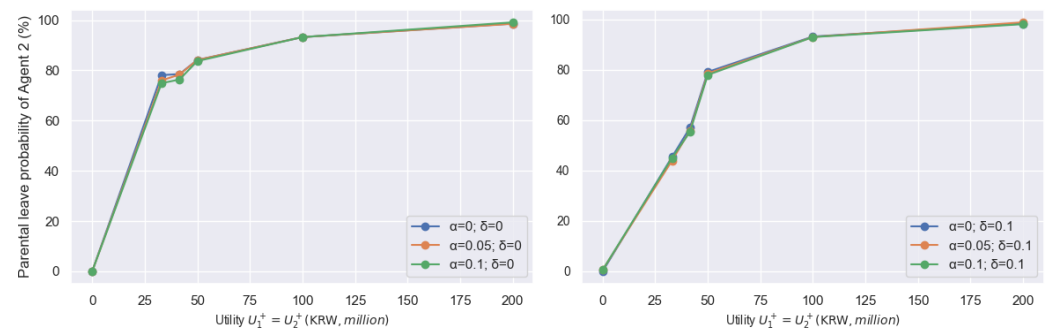


Figure 2. Comparison of the parental leave probability of agent 2 when $U_1^+ = U_2^+$, $\delta \in \{0, 0.1\}$, and α varies among $\{0, 0.05, 0.1\}$. Under symmetric utility conditions $U_1^+ = U_2^+$, the impact of the competitive penalty (α) on the leave probability is minimal.

5.2.2. Results Under Different Utility Values

When two employees have different perceived utilities for parental leave, the influence of the competitive career penalty (α) becomes more distinct. Two employees can show different parental leave usage patterns even in the same states because they perceive different marginal benefits, which can result in situations where one employee takes parental leave while the other does not. Decisions in such states reveal the influences of interpersonal factors on parental leave decisions.

Figure 3 illustrates how employee 2's parental leave behavior changes under different levels of α , δ , and U_2^+ when employee 1 does not take parental leave at all ($U_1^+ = 0$). In such settings, employee 2 is more likely to be affected by the competitive career penalty and tends to forgo parental leave as α increases from 0 to 0.1. For example, when $U_1^+ = 0$, $U_2^+ = 33$ M, and $\delta = 0$, employee 2 takes parental leave with a probability of 0.78 if $\alpha = 0$, but only 0.53 if $\alpha = 0.1$. The effects of α are further amplified when combined with the individual career penalty δ . Figure 3 shows that when $\delta = 0$, the parental leave probability of employee 2 is significantly affected by α if $U_2^+ = 33$ M. However, when $\delta = 0.1$, the effect of α becomes significant even when $U_2^+ = 50$ M. This implies that even when an employee perceives a relatively high utility from taking parental leave, the combined effects of competitive and individual career penalties can further suppress their willingness to take parental leave.

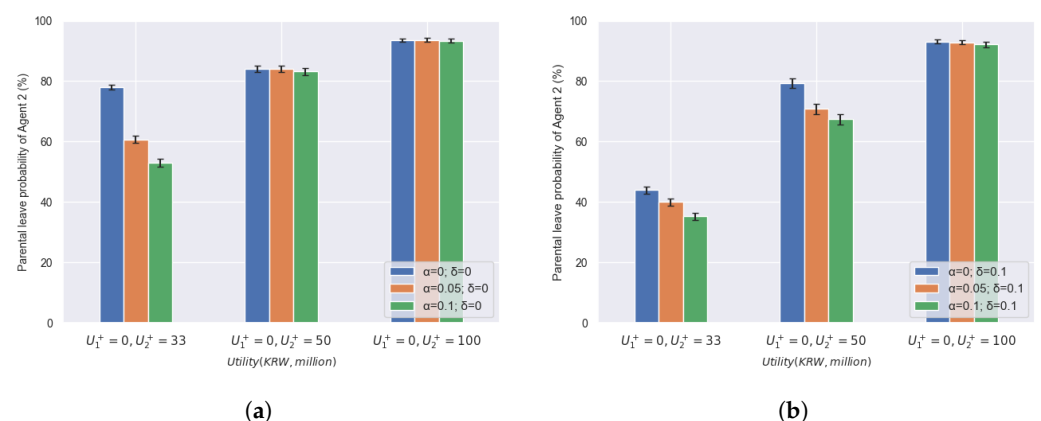


Figure 3. The parental leave probability of agent 2 when $U_1^+ = 0$, $U_2^+ \in \{33\text{ M}, 50\text{ M}, 100\text{ M}\}$. In asymmetric settings, a higher α significantly reduces leave uptake of agent 2 when U_2^+ is moderate and δ exists, but has a negligible effect when U_2^+ is very high. (a) Comparison of the parental leave probability of agent 2 when $\delta = 0$ and α varies among $\{0, 0.05, 0.1\}$. (b) Comparison of the parental leave probability of agent 2 when $\delta = 0.1$ and α varies among $\{0, 0.05, 0.1\}$.

Even when an employee's perceived utility value (an individual factor) and the extent of penalties after taking parental leave (organizational and community factors) are held constant, the probability of taking parental leave varies with the colleague's utility values, highlighting the influence of interpersonal factors. Figures 4 and 5 show how the parental leave probability of Agent 2 changes with the other employee's utility value (U_1^+) when the other influencing factors are held constant. Figures 4 and 5 consider the U_2^+ values of 33 M and 50 M, respectively.

In both figures, if $\alpha > 0$, which indicates the presence of interpersonal influences, the parental leave probability decreases as the colleague's perceived utility value decreases. For example, when $U_2^+ = 33$ M, $\alpha = 0.05$, and $\delta = 0.1$, the probability of parental leave is 0.44 if $U_1^+ = 33$ M, but drops to 0.40 if $U_1^+ = 0$. This implies that an employee may hesitate to take parental leave, even when their own willingness to take parental leave remains unchanged, if the colleague does not take parental leave. This occurs because the employee may lose long-term benefits even under the same perceived utility on parental leave, when the colleague does not take parental leave and the competitive career penalty is activated. Thus, the influence of the colleague's utility value is more pronounced when it is lower than the employee's utility value, as shown in Figures 4 and 5. Conversely, if the colleague's utility value is greater than that of the employee ($U_1^+ \geq U_2^+$), the parental leave probability of the employee is close to the value observed when $\alpha = 0$. As the colleague is more likely to take parental leave, the competitive career penalty is rarely triggered, and hence the employee behaves as though the competitive career penalty does not exist. When the competitive career penalty is absent, parental leave decisions appear independent of the utility of the colleague.

In summary, numerical experiments clarify the effects of various factors on parental leave decisions. Employees are more likely to take parental leave if they perceive the value of parental leave to be higher. Career penalties can discourage employees from taking parental leave. A higher individual career penalty decreases the parental leave probability. The competitive career penalty is effective when two employees have different perceived utility values. The influence of these two types of penalty is compounded; hence, the combination of the two career penalties further discourages employees from taking parental leave. Lastly, the colleague's perceived utility value affects the parental leave probability of the employee. When the colleague is less likely to take parental leave, even when other factors are held constant, the employee may hesitate to take parental leave due to concerns about disadvantages from the competitive penalty.

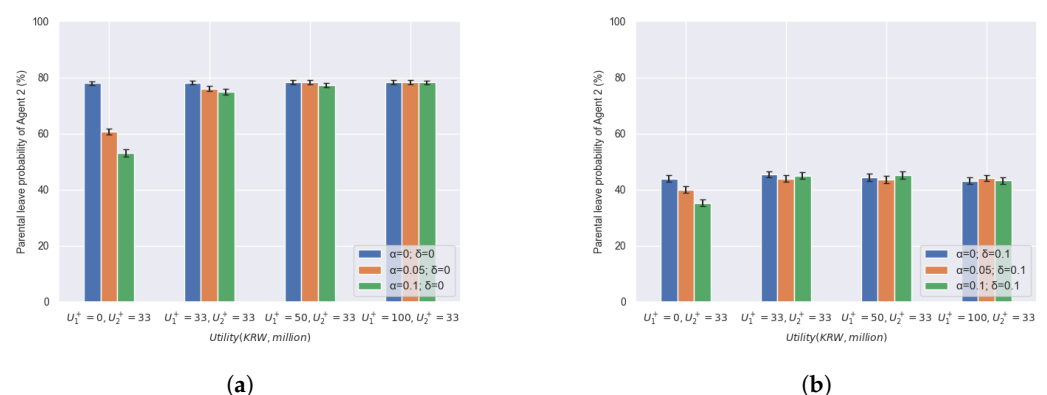


Figure 4. The parental leave probability of agent 2 when $U_2^+ = 33$ M, $U_1^+ \in \{0, 33$ M, 50 M, 100 M $\}$. Even when agent 2 has the same perceived utility, differences in agent 1's parental leave behavior affect agent 2's leave-taking probability. (a) Comparison of the parental leave probability of agent 2 when $\delta = 0$ and α varies among $\{0, 0.05, 0.1\}$. (b) Comparison of the parental leave probability of agent 2 when $\delta = 0.1$ and α varies among $\{0, 0.05, 0.1\}$.

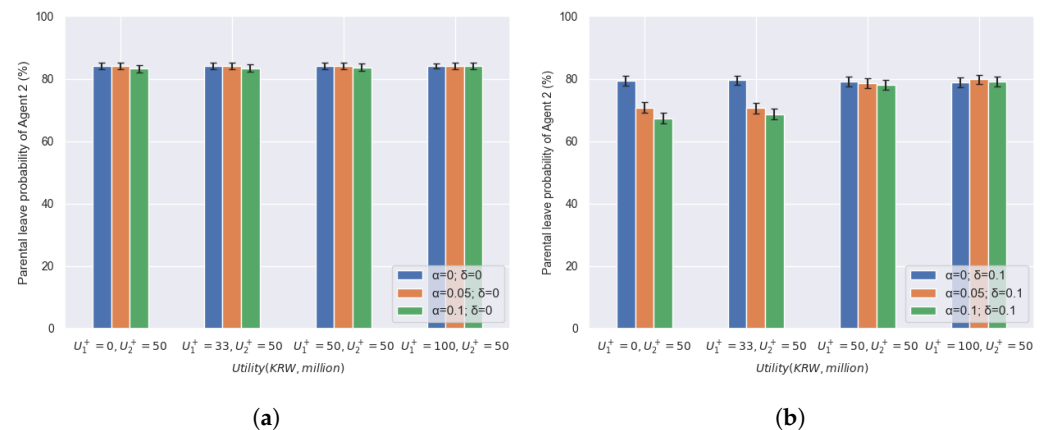


Figure 5. The parental leave probability of agent 2 when $U_2^+ = 50$ M, $U_1^+ \in \{0, 33 \text{ M}, 50 \text{ M}, 100 \text{ M}\}$. When agent 1 has a relatively low perceived utility (0 or 33 M), agent 2's parental leave probability decreases. (a) Comparison of the parental leave probability of agent 2 when $\delta = 0$ and α varies among $\{0, 0.05, 0.1\}$. (b) Comparison of the parental leave probability of agent 2 when $\delta = 0.1$ and α varies among $\{0, 0.05, 0.1\}$.

These findings underscore the importance of environmental as well as individual factors in parental leave decisions. Experimental results have repeatedly shown that the extent of career penalties should be reduced to encourage employees to take parental leave. Notably, the influence of a colleague's parental leave behavior is effective only when the colleague's willingness to take parental leave is lower than that of the affected employee and when the competitive career penalty exists. This suggests that negative signals regarding the use of parental leave are more influential than positive ones. Therefore, avoiding discrimination against employees who have taken parental leave may be more critical for fostering a parental-leave-friendly organizational culture than simply having employees who have a strong willingness to take parental leave.

In our numerical experiments, the time horizon, discount factor, and base promotion probability were fixed in order to isolate the effects of career-related penalties. However, the magnitude of these effects may vary depending on the specific values of these structural parameters. If the time horizon, which represents the total length of an employee's career, increases, the cumulative reward associated with successful promotion becomes larger. As a result, the impact of career-related penalties may become more pronounced. The discount factor reflects the extent to which employees value future rewards relative to immediate rewards. A lower discount factor places less weight on future outcomes and may therefore attenuate the influence of career-related penalties, which primarily affect future promotion opportunities. Similarly, a lower base promotion probability increases the perceived risk of promotion failure. Under such conditions, the marginal effect of career-related penalties may become stronger.

6. Conclusions

This study investigates the effects of multidimensional factors on parental leave decisions using an SG model and a multi-agent reinforcement learning algorithm (MARL). The SG model incorporates individual-, interpersonal-, organizational-, community-, and public policy-level factors on parental leave decisions in an integrated manner by considering perceived utility from parental leave, competitive and individual career penalties, and the subsidy rule. In contrast to previous studies that rely on empirical data, this study leverages the advantages of a mathematical modeling approach to thoroughly examine the effects of each decision factor. In particular, a novel game-theoretic framework is introduced to

capture employees' strategic parental leave decisions in a competitive work environment. Solutions to the proposed SG model are obtained through an extended Nash Q-learning algorithm. The proposed algorithm integrates backward iteration and optimistic initialization methods to exploit the structural properties of the proposed SG model, thereby improving the convergence speed and learning accuracy.

Numerical studies clarify the effects of career penalties and perceived utility values under various conditions. The experimental results show that both the individual career penalty and the competitive career penalty discourage employees from taking parental leave. The effects of these two types of penalties can be compounded, further reducing the parental leave probability. While an employee is more likely to take parental leave when the employee's perceived utility is sufficiently high, the employee tends to forgo parental leave if the colleague has a lower perceived utility value and the competitive career penalty is present. These findings suggest that negative perceptions of parental leave at both the individual and organizational levels are critical factors in reducing its usage.

Although this study identifies the effects of multidimensional factors on parental leave decisions through a mathematical model and simulation experiments, the proposed model and algorithm consider only two agents for simplicity. Moreover, our framework assumes one child and a one-year parental leave, whereas in reality, employees may have multiple children or take leave across several years, leading to cumulative or dynamic career effects. Future research could extend the current framework to N-player stochastic games by leveraging scalable multi-agent RL (e.g., mean-field Q-learning) to examine whether the interpersonal effects identified in this study persist in larger multi-agent settings. Additional extensions could incorporate alternative promotion mechanisms, such as rank-order tournaments or settings with limited promotion opportunities, as well as richer penalty structures. For example, proportion-based penalties could be considered, in which the extent of the competitive career penalty scales with the fraction of colleagues who do not take parental leave.

Author Contributions: Conceptualization, L.Z. and H.-R.L.; methodology, L.Z. and H.-R.L.; software, L.Z.; validation, L.Z.; writing—original draft preparation, L.Z.; writing—review and editing, L.Z. and H.-R.L.; visualization, L.Z.; supervision, H.-R.L.; project administration, H.-R.L.; funding acquisition, H.-R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by INHA UNIVERSITY Research Grant.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Appendix A.1. Single-Agent MDP Model

The single-agent MDP model is defined by specifying a state, an action and a reward function of the MDP model through an individual state s_t^i , an individual action a_t^i and an individual reward function $r^i(s_t, a_t)$ for agent i , respectively. The transition dynamics of state elements in the MDP model are identical to those in the SG model, except that the promotion probability $q(x^i, m^i, m^{-i})$ is set to $q(x^i, m^i, 0)$, which corresponds to the absence of competitive career penalty. Consequently, the single-agent MDP model and the multi-agent SG model are coupled through their promotion probability structure, as illustrated in Figure A1.

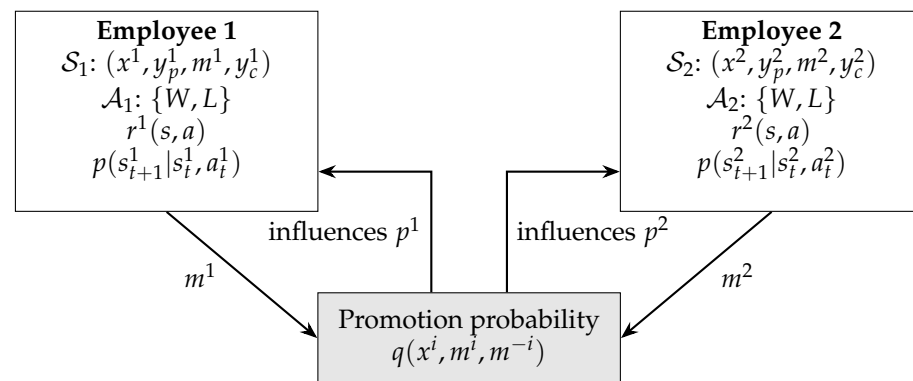


Figure A1. Strategic coupling mechanism of two-player MDPs.

Appendix A.2. Full Results of Experiments

Table A1. Mean parental leave probability(%) of two agents with 95% confidence interval when $U_1^+ = U_2^+$.

Parameters	Perceived Utility Level ($U_1^+ = U_2^+$)					
	0 M	33 M	41.5 M	50 M	100 M	200 M
$\alpha = 0$	0.00 $\pm$ 0.00	78.34 $\pm$ 0.79	78.76 $\pm$ 0.79	84.08 $\pm$ 1.08	93.22 $\pm$ 0.71	98.61 $\pm$ 0.41
$\delta = 0$	0.00 $\pm$ 0.00	78.12 $\pm$ 0.71	78.52 $\pm$ 0.71	84.15 $\pm$ 1.02	93.25 $\pm$ 0.66	98.62 $\pm$ 0.38
$\alpha = 0$	0.00 $\pm$ 0.00	45.93 $\pm$ 1.19	57.30 $\pm$ 1.22	78.85 $\pm$ 1.68	93.15 $\pm$ 0.73	98.34 $\pm$ 0.51
$\delta = 0.1$	0.01 $\pm$ 0.02	45.41 $\pm$ 1.07	57.00 $\pm$ 1.07	79.19 $\pm$ 1.57	93.25 $\pm$ 0.66	98.36 $\pm$ 0.46
$\alpha = 0$	0.00 $\pm$ 0.01	31.79 $\pm$ 1.31	40.74 $\pm$ 1.40	60.94 $\pm$ 1.61	91.64 $\pm$ 1.17	98.37 $\pm$ 0.49
$\delta = 0.2$	0.01 $\pm$ 0.02	31.10 $\pm$ 1.23	40.55 $\pm$ 1.36	61.40 $\pm$ 1.61	91.85 $\pm$ 1.05	98.44 $\pm$ 0.44
$\alpha = 0.05$	0.00 $\pm$ 0.01	76.31 $\pm$ 0.97	78.56 $\pm$ 0.80	84.08 $\pm$ 1.08	93.22 $\pm$ 0.71	98.54 $\pm$ 0.43
$\delta = 0$	0.01 $\pm$ 0.02	76.00 $\pm$ 0.88	78.34 $\pm$ 0.71	84.15 $\pm$ 1.02	93.25 $\pm$ 0.66	98.55 $\pm$ 0.40
$\alpha = 0.1$	0.02 $\pm$ 0.02	75.09 $\pm$ 1.08	76.58 $\pm$ 1.02	83.54 $\pm$ 1.18	93.22 $\pm$ 0.71	99.15 $\pm$ 0.24
$\delta = 0$	0.02 $\pm$ 0.02	74.89 $\pm$ 0.99	76.37 $\pm$ 0.91	83.70 $\pm$ 1.09	93.25 $\pm$ 0.66	99.13 $\pm$ 0.24
$\alpha = 0.05$	0.50 $\pm$ 0.16	44.34 $\pm$ 1.31	56.12 $\pm$ 1.30	78.24 $\pm$ 1.73	92.95 $\pm$ 0.80	98.84 $\pm$ 0.36
$\delta = 0.1$	0.58 $\pm$ 0.16	43.97 $\pm$ 1.23	55.85 $\pm$ 1.14	78.58 $\pm$ 1.58	93.07 $\pm$ 0.71	98.88 $\pm$ 0.33
$\alpha = 0.1$	0.50 $\pm$ 0.16	45.08 $\pm$ 1.28	55.60 $\pm$ 1.31	77.71 $\pm$ 1.78	92.91 $\pm$ 0.80	98.14 $\pm$ 0.56
$\delta = 0.1$	0.56 $\pm$ 0.16	45.04 $\pm$ 1.23	55.35 $\pm$ 1.15	78.00 $\pm$ 1.63	93.03 $\pm$ 0.71	98.17 $\pm$ 0.51

Note: For each cell, upper and lower values represent Agent 1 and Agent 2, respectively.

Table A2. Mean parental leave probability (%) of two agents with 95% confidence interval when $U_1^+ \neq U_2^+$.

Parameters	Perceived Utility Combination (U_1^+, U_2^+)					
	0 M 33 M	0 M 50 M	0 M 100 M	33 M 50 M	33 M 100 M	100 M 50 M
$\alpha = 0$	0.26 $\pm$ 0.07	0.17 $\pm$ 0.06	0.06 $\pm$ 0.04	78.21 $\pm$ 0.78	78.36 $\pm$ 0.80	93.22 $\pm$ 0.71
$\delta = 0$	78.00 $\pm$ 0.69	84.15 $\pm$ 1.02	93.55 $\pm$ 0.59	84.15 $\pm$ 1.02	93.45 $\pm$ 0.62	84.12 $\pm$ 0.88
$\alpha = 0.05$	0.47 $\pm$ 0.11	0.01 $\pm$ 0.01	0.01 $\pm$ 0.01	78.31 $\pm$ 0.80	78.20 $\pm$ 0.78	93.22 $\pm$ 0.71
$\delta = 0$	60.78 $\pm$ 1.10	84.15 $\pm$ 1.02	93.65 $\pm$ 0.58	84.15 $\pm$ 1.02	93.25 $\pm$ 0.66	84.14 $\pm$ 1.00
$\alpha = 0.1$	0.01 $\pm$ 0.01	0.06 $\pm$ 0.03	0.14 $\pm$ 0.05	77.27 $\pm$ 0.89	78.18 $\pm$ 0.77	93.27 $\pm$ 0.70
$\delta = 0$	53.09 $\pm$ 1.30	83.20 $\pm$ 1.14	93.37 $\pm$ 0.63	83.42 $\pm$ 1.10	93.31 $\pm$ 0.65	84.15 $\pm$ 0.99
$\alpha = 0$	0.06 $\pm$ 0.04	0.03 $\pm$ 0.02	0.15 $\pm$ 0.06	44.47 $\pm$ 1.28	43.22 $\pm$ 1.18	93.47 $\pm$ 0.65
$\delta = 0.1$	43.99 $\pm$ 1.22	79.42 $\pm$ 1.56	93.08 $\pm$ 0.71	79.51 $\pm$ 1.49	93.29 $\pm$ 0.65	78.83 $\pm$ 1.48
$\alpha = 0.05$	0.06 $\pm$ 0.04	0.40 $\pm$ 1.14	0.63 $\pm$ 0.20	43.61 $\pm$ 1.25	44.12 $\pm$ 1.14	92.51 $\pm$ 0.92
$\delta = 0.1$	40.06 $\pm$ 1.22	70.80 $\pm$ 1.65	92.89 $\pm$ 0.77	70.66 $\pm$ 1.70	93.25 $\pm$ 0.67	79.83 $\pm$ 1.49
$\alpha = 0.1$	0.72 $\pm$ 0.16	0.12 $\pm$ 0.05	0.18 $\pm$ 0.06	45.27 $\pm$ 1.28	43.29 $\pm$ 1.17	92.40 $\pm$ 0.94
$\delta = 0.1$	35.32 $\pm$ 1.20	67.39 $\pm$ 1.76	92.18 $\pm$ 0.95	68.71 $\pm$ 1.67	92.44 $\pm$ 0.89	79.16 $\pm$ 1.52

Note: For each cell, upper and lower values represent Agent 1 and Agent 2, respectively.

Appendix A.3. Pseudo-Code of Nash Q-Learning

Algorithm A1 Nash Q-Learning

```

1: Initialize:  $t \leftarrow 0$ ; obtain initial state  $s_0$ 
2: Let the learning agent be indexed by  $i$ 
3: for all  $s \in S, a^i \in A_i, i = 1, \dots, N$  do
4:    $Q^i(s, a_1^1, \dots, a_t^N) \leftarrow 0$ 
5: end for
6: while not converged do
7:   Choose action  $a_t^1, \dots, a_t^N$ 
8:   Observe  $r_t^1, \dots, r_t^N$  and  $s_{t+1}$ 
9:   for  $i = 1$  to  $N$  do
10:     $Q^i(s_t, a_t^1, \dots, a_t^N) \leftarrow (1 - l_t)Q^i(s_t, a_t^1, \dots, a_t^N) + l_t[r_t^i + \gamma \text{Nash}Q^i(s_{t+1})]$ 
11:    where  $l_t \in (0, 1)$  is the learning rate at the  $t^{\text{th}}$  updates, and  $\text{Nash}Q^i(s_{t+1})$  is the Nash
    q-value
12:   end for
13:    $t \leftarrow t + 1$ 
14: end while

```

Appendix A.4. Ablation Results

Table A3 presents the ablation results evaluating the effectiveness of optimistic initialization and backward iteration on convergence speed. The proposed method converges approximately three times faster than vanilla Nash-Q learning. Convergence is defined as the point at which the maximum relative change in Q-values between consecutive iterations is less than 0.001% of the Q-values and is maintained for 200 consecutive iterations. Due to the substantial computational cost of the ablation analysis, we report results for a representative parameter setting ($\alpha = 0.1, \delta = 0, U_1^+ = 4150 \text{ M}, U_2^+ = 4150 \text{ M}$). Similar convergence improvements were consistently observed across other parameter settings during the development of the algorithm.

Table A3. Average number of update episodes required for convergence across all states of interest.

Setting	Convergence Speed
vanilla Nash Q-learning	3668
Nash Q-learning + backward iteration	3452
Nash Q-learning + optimistic init	1116
Nash Q-learning + optimistic init + backward iteration	1013

Appendix B. Code Availability

The full implementation used in this study is publicly available at https://github.com/ggstk/ParentalLeave_RL (accessed on 1 February 2026). The repository includes the environment, Nash Q-learning implementation, hyperparameter configurations, and scripts to reproduce all experiments reported in the paper.

References

- Haas, L.; Hwang, C.P. The impact of taking parental leave on fathers' participation in childcare and relationships with children: Lessons from Sweden. *Community Work. Fam.* **2008**, *11*, 85–104. [\[CrossRef\]](#)
- O'Brien, M. Fathers, parental leave policies, and infant quality of life: International perspectives and policy impact. *Ann. Am. Acad. Political Soc. Sci.* **2009**, *624*, 190–213. [\[CrossRef\]](#)
- Bünning, M. What happens after the 'daddy months'? Fathers' involvement in paid work, childcare, and housework after taking parental leave in Germany. *Eur. Sociol. Rev.* **2015**, *31*, 738–748. [\[CrossRef\]](#)
- Heshmati, A.; Honkaniemi, H.; Juárez, S.P. The effect of parental leave on parents' mental health: A systematic review. *Lancet Public Health* **2023**, *8*, e57–e75. [\[CrossRef\]](#) [\[PubMed\]](#)
- Burt, A.; Bezruchka, S. Population health and paid parental leave: What the United States can learn from two decades of research. *Healthcare* **2016**, *4*, 30. [\[CrossRef\]](#)

6. Tanaka, S. Parental leave and child health across OECD countries. *Econ. J.* **2005**, *115*, F7–F28. [\[CrossRef\]](#)
7. Baum, C.L.; Ruhm, C.J. The effects of paid family leave in California on labor market outcomes. *J. Policy Anal. Manag.* **2016**, *35*, 333–356. [\[CrossRef\]](#)
8. Thévenon, O.; Solaz, A. *Labour Market Effects of Parental Leave Policies in OECD Countries*; OECD Social, Employment and Migration Working Papers; OECD: Paris, France, 2013.
9. Doucet, A.; McKay, L. Fathering, parental leave, impacts, and gender equality: What/how are we measuring? *Int. J. Sociol. Soc. Policy* **2020**, *40*, 441–463. [\[CrossRef\]](#)
10. Bastani, S.; Blumkin, T.; Micheletto, L. The welfare-enhancing role of parental leave mandates. *J. Law Econ. Organ.* **2019**, *35*, 77–126. [\[CrossRef\]](#)
11. Carlsson, M.; Reshid, A.A. Co-worker peer effects on parental leave take-up. *Scand. J. Econ.* **2022**, *124*, 930–957.
12. Akerlof, G.A.; Kranton, R.E. Economics and identity. *Q. J. Econ.* **2000**, *115*, 715–753. [\[CrossRef\]](#)
13. Haas, L.; Allard, K.; Hwang, P. The impact of organizational culture on men’s use of parental leave in Sweden. *Community Work. Fam.* **2002**, *5*, 319–342. [\[CrossRef\]](#)
14. Petts, R.J.; Mize, T.D.; Kaufman, G. Organizational policies, workplace culture, and perceived job commitment of mothers and fathers who take parental leave. *Soc. Sci. Res.* **2022**, *103*, 102651. [\[CrossRef\]](#) [\[PubMed\]](#)
15. McKay, L.; Doucet, A. “Without taking away her leave”: A canadian case study of couples’ decisions on fathers’ use of paid parental leave. *Fathering* **2010**, *8*, 300. [\[CrossRef\]](#)
16. Romero-Balsas, P.; Muntanyola-Saura, D.; Rogero-García, J. Decision-making factors within paternity and parental leaves: Why Spanish fathers take time off from work. *Gender Work. Organ.* **2013**, *20*, 678–691. [\[CrossRef\]](#)
17. Lee, Y. ‘Undoing gender’ or selection effects?: Fathers’ uptake of leave and involvement in housework and childcare in South Korea. *J. Fam. Stud.* **2023**, *29*, 2430–2458. [\[CrossRef\]](#)
18. Lee, Y. Norms about childcare, working hours, and fathers’ uptake of parental leave in South Korea. *Community Work. Fam.* **2023**, *26*, 466–491. [\[CrossRef\]](#)
19. McLeroy, K.R.; Bibeau, D.; Steckler, A.; Glanz, K. An ecological perspective on health promotion programs. *Health Educ. Q.* **1988**, *15*, 351–377. [\[CrossRef\]](#)
20. Hu, J.; Wellman, M.P. Nash Q-learning for general-sum stochastic games. *J. Mach. Learn. Res.* **2003**, *4*, 1039–1069.
21. Ding, Z.W.; Zheng, G.Z.; Cai, C.R.; Cai, W.R.; Chen, L.; Zhang, J.Q.; Wang, X.M. Emergence of cooperation in two-agent repeated games with reinforcement learning. *Chaos Solitons Fractals* **2023**, *175*, 114032. [\[CrossRef\]](#)
22. Leslie, D.S.; Collins, E.J. Individual Q-learning in normal form games. *SIAM J. Control Optim.* **2005**, *44*, 495–514. [\[CrossRef\]](#)
23. Kaufman, G. Barriers to equality: Why British fathers do not use parental leave. *Community Work. Fam.* **2018**, *21*, 310–325. [\[CrossRef\]](#)
24. Jørgensen, T.H.; Søgaaard, J.E. Welfare Reforms and the Division of Parental Leave. 2021. Available online: <https://ssrn.com/abstract=3831467> (accessed on 1 February 2026).
25. Meil, G.; García Sainz, C.; Luque, M.; Ayuso, L. *El Impacto de los Permisos Parentales en la Carrera Profesional*; Universidad Autónoma de Madrid: Madrid, Spain, 2007.
26. Haas, L.; Hwang, P.O. Company Culture and Men’s Usage of Family Leave Benefits in Sweden. *Fam. Relat.* **1995**, *44*, 28–36.
27. Fried, M. *Taking Time*; Temple University Press: Philadelphia, PA, USA, 1998; Volume 9.
28. Margolis, R.; Hou, F.; Haan, M.; Holm, A. Use of parental benefits by family income in Canada: Two policy changes. *J. Marriage Fam.* **2019**, *81*, 450–467.
29. Schmeer, J.A.; Reitman, F. The interrupted managerial career path: A longitudinal study of MBAs. *J. Vocat. Behav.* **1997**, *51*, 411–434. [\[CrossRef\]](#)
30. Judiesch, M.K.; Lyness, K.S. Left behind? The impact of leaves of absence on managers’ career success. *Acad. Manag. J.* **1999**, *42*, 641–651.
31. Evertsson, M. Parental leave and careers: Women’s and men’s wages after parental leave in Sweden. *Adv. Life Course Res.* **2016**, *29*, 26–40. [\[CrossRef\]](#)
32. Tô, L.T. The Signaling Role of Parental Leave. Ph.D. Thesis, Harvard University, Cambridge, MA, USA, 2018.
33. Chatterji, P.; Markowitz, S. *Family Leave After Childbirth and the Health of New Mothers*; Working Paper 14156; National Bureau of Economic Research: Cambridge, MA, USA, 2008. [\[CrossRef\]](#)
34. Van Niel, M.S.; Bhatia, R.; Riano, N.S.; De Faria, L.; Catapano-Friedman, L.; Ravven, S.; Weissman, B.; Nzodom, C.; Alexander, A.; Budde, K.; et al. The impact of paid maternity leave on the mental and physical health of mothers and children: A review of the literature and policy implications. *Harv. Rev. Psychiatry* **2020**, *28*, 113–126. [\[CrossRef\]](#)
35. Ruhm, C.J. Parental leave and child health. *J. Health Econ.* **2000**, *19*, 931–960.
36. Danzer, N.; Halla, M.; Schneeweis, N.; Zweimüller, M. Parental leave, (in)formal childcare, and long-term child outcomes. *J. Hum. Resour.* **2022**, *57*, 1826–1884. [\[CrossRef\]](#)
37. Huber, K. Changes in parental leave and young children’s non-cognitive skills. *Rev. Econ. Househ.* **2019**, *17*, 89–119. [\[CrossRef\]](#)

38. Liu, Q.; Skans, O.N. The Duration of Paid Parental Leave and Children's Scholastic Performance. *B.E. J. Econ. Anal. Policy* **2010**, *10*, 3. [[CrossRef](#)]
39. Nick, J.M.; Sahin, S.; Roberts, L.R.; Hatton, A.; Cafferky, B. Effect of paternity leave or fathers' parental leave on infant health: A systematic review protocol. *JBMEvid. Synth.* **2025**, *23*, 792–800. [[CrossRef](#)] [[PubMed](#)]
40. del Carmen Huerta, M.; Adema, W.; Baxter, J.; Han, W.J.; Lausten, M.; Lee, R.; Waldfogel, J. *Fathers' Leave, Fathers' Involvement and Child Development: Are They Related? Evidence from Four OECD Countries*; OECD: Paris, France, 2013.
41. Ekberg, J.; Eriksson, R.; Friebe, G. Parental leave—A policy evaluation of the Swedish “Daddy-Month” reform. *J. Public Econ.* **2013**, *97*, 131–143. [[CrossRef](#)]
42. Patnaik, A. Reserving time for daddy: The consequences of fathers' quotas. *J. Labor Econ.* **2019**, *37*, 1009–1059. [[CrossRef](#)]
43. Castro-García, C.; Pazos-Moran, M. Parental leave policy and gender equality in Europe. *Fem. Econ.* **2016**, *22*, 51–73. [[CrossRef](#)]
44. Shapley, L.S. Stochastic Games. *Proc. Natl. Acad. Sci. USA* **1953**, *39*, 1095–1100. [[CrossRef](#)]
45. Lemke, C.E.; Howson, J.T., Jr. Equilibrium points of bimatrix games. *J. Soc. Ind. Appl. Math.* **1964**, *12*, 413–423. [[CrossRef](#)]
46. Ministry of Employment and Labor of Korea. *Wage and Job Information System*; Ministry of Employment and Labor of Korea: Sejong-si, Republic of Korea, 2024.
47. Joo, M. *Wage System and the Workforce Management Survey Report of 2023*; Technical Report; Ministry of Employment and Labor of Korea: Sejong-si, Republic of Korea, 2023.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.