HAL
open science

Entangled and correlated photon mixed strategy for social decision making

Shion Maeda, Nicolas Chauvet, Hayato Saigo, Hirokazu Hori, Guillaume Bachelier, Serge Huant, Makoto Naruse

To cite this version:

Shion Maeda, Nicolas Chauvet, Hayato Saigo, Hirokazu Hori, Guillaume Bachelier, et al.. Entangled and correlated photon mixed strategy for social decision making. Scientific Reports, 2021, 11, pp.4832. 10.1038/s41598-021-84199-5 . hal-03376291

HAL Id: hal-03376291

https://hal.science/hal-03376291v1

Submitted on 13 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

scientific reports

OPEN

Entangled and correlated photon mixed strategy for social decision making

Shion Maeda[1]✉, Nicolas Chauvet[1,2], Hayato Saigo[3], Hirokazu Hori[4], Guillaume Bachelier[5], Serge Huant[5] & Makoto Naruse[1,2]✉

Collective decision making is important for maximizing total benefits while preserving equality among individuals in the competitive multi-armed bandit (CMAB) problem, wherein multiple players try to gain higher rewards from multiple slot machines. The CMAB problem represents an essential aspect of applications such as resource management in social infrastructure. In a previous study, we theoretically and experimentally demonstrated that entangled photons can physically resolve the difficulty of the CMAB problem. This decision-making strategy completely avoids decision conflicts while ensuring equality. However, decision conflicts can sometimes be beneficial if they yield greater rewards than non-conflicting decisions, indicating that greedy actions may provide positive effects depending on the given environment. In this study, we demonstrate a mixed strategy of entangled- and correlated-photon-based decision-making so that total rewards can be enhanced when compared to the entangled-photon-only decision strategy. We show that an optimal mixture of entangled- and correlated-photon-based strategies exists depending on the dynamics of the reward environment as well as the difficulty of the given problem. This study paves the way for utilizing both quantum and classical aspects of photons in a mixed manner for decision making and provides yet another example of the supremacy of mixed strategies known in game theory, especially in evolutionary game theory.

Optics and photonics are expected to play key roles in accommodating the massive requirements of future intelligent information systems[1]. Recent developments in the field include photonic reservoir computing for time series predictions[2–4], on-chip lightwave circuits for photonic learning functions[5,6], fibre optic neuromorphic systems[7,8], among others. Decision making is another important research topic in which decisions have to be made autonomously in dynamically changing uncertain environments[9,10]. Furthermore, collective decision making involving multiple players becomes a critical issue in the management of social utilities[11]. Recently, photonic approaches to decision-making problems have been intensively studied using single photons[12,13], chaotic lasers[14,15], and entangled photons[16]. Related topics are discussed in recent review articles[1,17,18].

The multi-armed bandit (MAB) problem describes some of the fundamental issues associated with decision making. The objective of the MAB problem is to maximize the player rewards from multiple slot machines with an initially unknown hit probability. Here, spending a lot of time gathering new information can be costly, while hasty decisions lead to missing out on good choices. This issue is known as the exploration–exploitation dilemma[19]. The physical attributes of photons have been successfully utilized to solve such MAB problems[20]. The MAB problem becomes even more complicated when multiple players come into play. As individual players seek to maximize their rewards, they will choose the best machine, which leads to a decision conflict. Many relevant issues in real-life situations, ranging from congestion in information networks and traffic jams on roads to hoarding of goods are caused by many players choosing the same decision[9,10], suggesting the importance of collective decision making. Such a problem that deals with multiple players and slot machines is called the competitive multi-armed bandit (CMAB) problem[16].

In previous work concerning the two-player, two-armed bandit problem, we theoretically and experimentally demonstrated collective decision making by using polarization-entangled photon pairs so that decision

[1]Department of Mathematical Engineering and Information Physics, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. [2]Department of Information Physics and Computing, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. [3]Nagahama Institute of Bio-Science and Technology, 1266 Tamura, Nagahama, Shiga 526-0829, Japan. [4]Interdisciplinary Graduate School, University of Yamanashi, Takeda, Kofu, Yamanashi 400-8510, Japan. [5]Univ. Grenoble Alpes, CNRS, Institut Néel, 38000 Grenoble, France. ✉email: maeda-shion4141@g.ecc.u-tokyo.ac.jp; makoto_naruse@ipc.i.u-tokyo.ac.jp

conflicts are avoided and maximum total reward is accomplished while ensuring equality[16]. More recently, we have theoretically derived optimal quantum states that provide the maximum total reward, while preserving equality, with respect to more than three players on two-armed bandit problems[21]. We also showed that classical photons, in the sense of not-entangled states such as single photons and correlated photon pairs, cannot resolve decision conflicts[16]. In these studies, the reward dispensed from a slot machine is constant at a time. Hence, in the event of a decision conflict, the individual player's reward is reduced by the number of overlaps, resulting in a reduction in the total reward.

However, depending on the given environmental conditions, decision conflicts can also provide a greater total reward. For example, if the individual reward is *not* reduced from a particular slot machine even in the case of conflicts, choosing the same slot machine (decision conflict) yields a higher total reward than choosing different slot machines (non-conflicting decision). Indeed, we can observe similar real-life scenarios, for example, in the form of enhanced services or resource availability, such as computing power and reduced sales prices which are offered during a limited amount of time. Similarly, the notion of critical mass that has to be reached for an activity to be sustainable reflects the non-decreasing rewards with conflicting decisions.

In this study, to accommodate the aforementioned changes in the environmental conditions and maximize total rewards, we propose and demonstrate a mixed strategy of utilizing entangled photons and classical photons (specifically, polarization-entangled photon pairs and polarization-correlated photon pairs) to find the optimal solution of 2-player, 2-armed bandit problems. While utilizing entangled photons, which guarantee non-conflicted and fully equal decisions, each player accumulates information about the reward environment. When recognizing that the conflicted choice provides greater rewards, we utilize correlated photons to fully exploit the reward from the environment. We show that an optimal mixture of entangled and correlated photons exists depending on the dynamics of the reward environment as well as the difficulty of finding the higher reward probability machine. Although the following discussion is restricted to 2-player, 2-choice problems, the present study captures the essential aspects of entangled and classical-photon mixed strategies that can be extended for solving more generalized problems.

Results

System architecture. We consider two players (Players 1 and 2), each of whom chooses one of two slot machines (Machines A and B) with the intention of maximizing the total reward or the summation of the reward of each player. The reward probabilities of Machines A and B are denoted as P_A and P_B, respectively. Although the present study examines the properties of entangled and classical photon states theoretically and numerically, it assumes technologically feasible experimental optical systems to generate photon pairs by spontaneous parametric down conversion (SPDC), as schematically represented in Fig. 1a, which is essentially the same as the experimental setup proposed in our previous study[16]. The photon pair generation is based on a standard Sagnac loop architecture[22] to induce SPDC. The signal and idler photons correspond to the decisions of Players 1 and 2, respectively. The signal photon goes through a half-wave plate (HW_1) followed by a polarization beam splitter PBS (PBS_1). If the photon is detected by the photodetector corresponding to the horizontally polarized light (PD1), the decision of Player 1 is to choose Machine A, whereas if the photon is detected by the photodetector corresponding to the vertically polarized light (PD2), then the decision of Player 1 is to choose Machine B. Similarly, the decision of player 2 is determined by the detection of the idler photon by PD3 or PD4, which corresponds to the decisions of selecting Machines A and B, respectively.

We introduce several notations to describe the system. The input photon state for the decision of Player i ($i = 1,\ 2$) is denoted as $|\theta_i\rangle$, where θ_i is the linear polarization angle. The roles of HW_i and PBS_i are given by

$$HW_i|\theta_i\rangle = \left|2\theta_{HW_i} - \theta_i\right\rangle \tag{1}$$

and

$$PBS_i\left|2\theta_{HW_i} - \theta_i\right\rangle = \cos(2\theta_{HW_i} - \theta_i)|H_i\rangle + \sin(2\theta_{HW_i} - \theta_i)|V_i\rangle, \tag{2}$$

where $|H_i\rangle$ and $|V_i\rangle$ indicate photon states with horizontal and vertical polarizations propagating in orthogonal directions beyond PBS_i[23]. One strategy for realizing collective decision making is to link the decisions of Players 1 and 2 by introducing correlations among the decisions at the level of photon states. Here, we consider polarization-orthogonal photon pairs denoted by, $|\theta_1, \theta_2\rangle$ where.

$$\theta_2 = \theta_1 + \pi/2, \tag{3}$$

as input photon states to the two players. In practice, we can fix θ_i ($i = 1, 2$) by controlling the polarizers and half/quarter waveplates in the path of the excitation laser (respectively denoted by P, HW_E, and QW_E in Fig. 1a). Let us set $\theta_1 = 0$ and $\theta_2 = \pi/2$, for the sake of simplicity. The probability of observing photons at PD1 and PD3 (meaning that both players chose Machine A) and at PD2 and PD4 (both players select Machine B) are represented by the following equations.

$$P_C(\mathrm{A},\mathrm{A}) = \cos^2\left(2\theta_{\mathrm{HW}_1}\right)\cos^2\left(2\theta_{\mathrm{HW}_2} - \frac{\pi}{2}\right) \tag{4}$$

and

$$P_C(\mathrm{B},\mathrm{B}) = \sin^2\left(2\theta_{\mathrm{HW}_1}\right)\sin^2\left(2\theta_{\mathrm{HW}_2} - \frac{\pi}{2}\right). \tag{5}$$

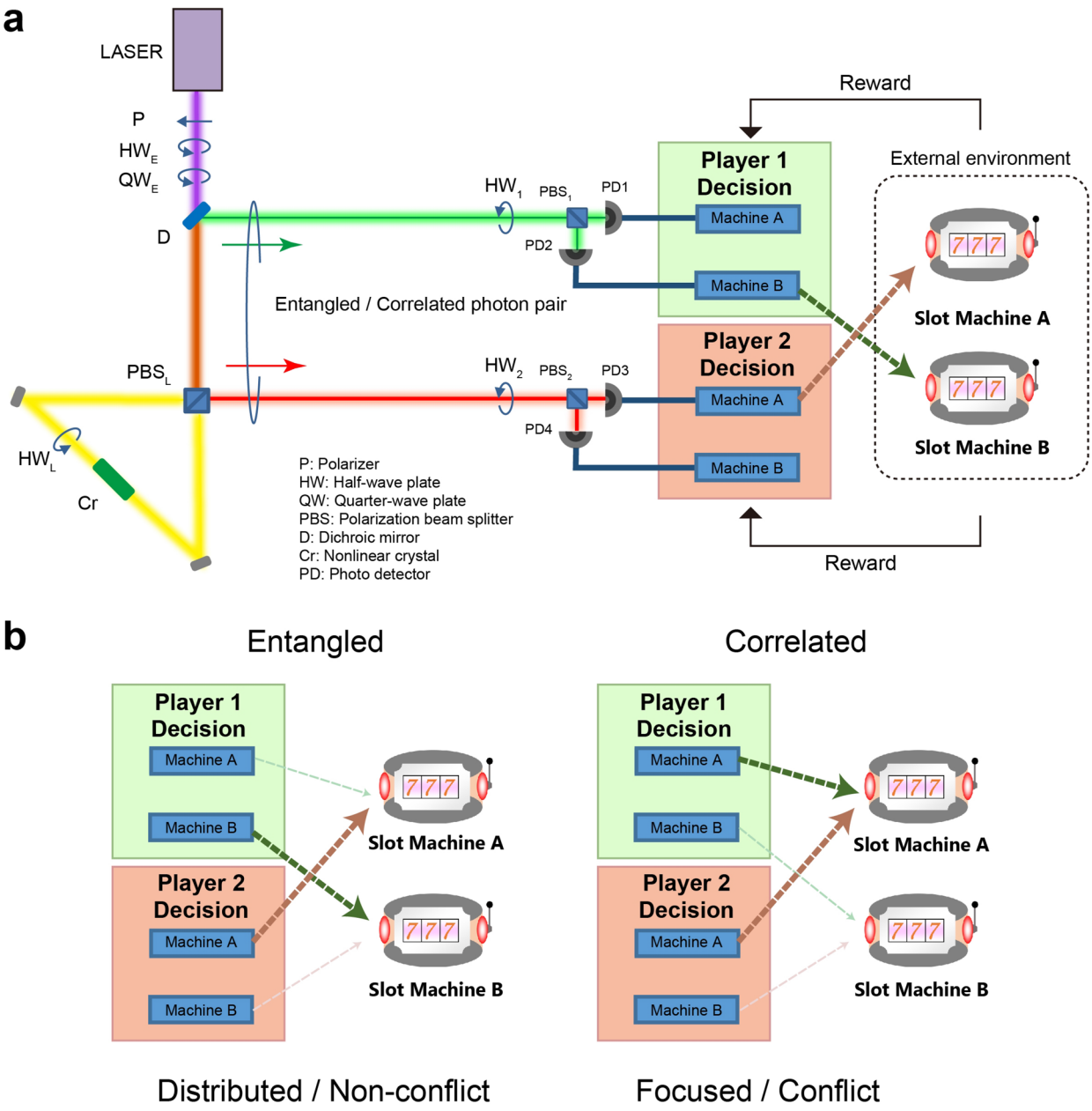

Figure 1. System architecture for entangled and correlated-photon mixed strategy for collective decision making. (**a**) Schematic of the optical system configuration consisting of photon pair (either entangled or correlated) generation and photon detection systems that directly provide the decisions of the two players (Player 1 and 2) to select either of the two slot machines (Machine A or Machine B) in the external environment. (**b**) With entangled photons, the decisions of the players are never in conflict. With correlated photons, both players take the same decision by properly choosing the half waveplate angle (see text for details). The elements of (**a**) are adapted from Chauvet et al.[16]. Copyright 2019 Author(s), licensed under a Creative Commons Attribution 4.0 License.

By letting $\theta_{HW_1} = 0$ and $\theta_{HW_2} = \pi/2$ or their $N\pi$ angle-shifted equivalents (where N is an integer) in Eq. (4), P(A, A) becomes unity, indicating that both players always chose Machine A, which is schematically illustrated in Fig. 1b. Similarly, P(B,B) becomes unity when $\theta_{HW_1} = \pi/2$ and $\theta_{HW_2} = 0$ or their $N\pi$ angle-shifted equivalents in Eq. (5). That is, with polarization orthogonal photon pairs, both players can choose the same intended machine with appropriate half-wave plate settings. Meanwhile, the probability of observing photons at PD1 and PD4 is given by

$$P_C(\mathrm{A},\mathrm{B}) = \cos^2\left(2\theta_{HW_1}\right)\sin^2\left(2\theta_{HW_2} - \frac{\pi}{2}\right) \tag{6}$$

which becomes unity when $\theta_{HW_1} = 0$ and $\theta_{HW_2} = 0$ or their $N\pi$ phase-shifted equivalent. P_C(A,B) = 1 implies that Player 1 always chooses Machine A while Player 2 always selects Machine B. There is indeed no decision

conflict in this case. However, equality is severely deteriorated; for example, when Machine A owns a higher reward probability than Machine B, Player 1 earns greater rewards than Player 2. More details can be found in Ref.[16]. In the discussion of mixed strategy below, such fixed choices prevent the players from autonomously realizing which machine is actually dispensing higher rewards.

To overcome this issue, we utilize a coherent superposition of states corresponding to the entangled states. Here, we consider the maximally entangled singlet photon state given by

$$\frac{1}{\sqrt{2}}(|\theta_1,\theta_2\rangle - |\theta_2,\theta_1\rangle), \tag{7}$$

where θ_1 and θ_2 are orthogonal to each other, as specified in Eq. (3). The maximally entangled photons are usually represented in the form $\frac{1}{\sqrt{2}}(|HV\rangle - |VH\rangle)$. A different notation is used in Eq. (7) to maintain consistent notations with the aforementioned polarization-correlated photons ($|\theta_1,\theta_2\rangle$) and to clearly present the role of half-wave plates in the following discussion. Considering the probability amplitude originating from the second term in Eq. (7), the probabilities of the two players' decisions are given by[16]:

$$P_{\mathrm{E}}(A,A) = P_{\mathrm{E}}(B,B) = \frac{1}{2}\sin^2\left[2\left(\theta_{\mathrm{HW}_1} - \theta_{\mathrm{HW}_2}\right)\right], \tag{8}$$

$$P_{\mathrm{E}}(A,B) = P_{\mathrm{E}}(B,A) = \frac{1}{2}\cos^2\left[2\left(\theta_{\mathrm{HW}_1} - \theta_{\mathrm{HW}_2}\right)\right], \tag{9}$$

which means that if $\theta_{\mathrm{HW}_1} = \theta_{\mathrm{HW}_2}$ is satisfied, the non-conflict probability is always unity and equality is ensured by Eq. (9). This also means that the conflict probability is always zero in Eq. (8) regardless of the values of θ_i and θ_{HW_i}. That is, both players randomly but equally select Machine A or B, but conflict never happens. Such collective decision-making is schematically illustrated in Fig. 1b. If θ_{HW_1} and θ_{HW_2} are orthogonally arranged with each other, for example, $\theta_{\mathrm{HW}_1} = \theta_{\mathrm{HW}_2} + \pi/2$, the relationship is completely reversed: decision conflict is always induced with equal probability at Machine A and Machine B. Although such an orthogonally arranged configuration is another interesting aspect of entangled-photon states given by Eq. (7), it is not exploited in the following discussion for simplicity. One remark is that, while such an orthogonal arrangement of θ_{HW_1} and θ_{HW_2} provides conflicted decisions, the machine to be chosen cannot be specified to the intended one as opposed to polarization correlated photon pairs discussed earlier; this aspect does not meet the mixed strategy discussed shortly below.

When two players make the same decision, the reward for each player is usually divided into two halves, as schematically illustrated in Fig. 2a. Therefore, from the viewpoint of maximizing the total reward, when a player chooses the best slot machine, the other player should select the other machine. Hence, entangled-photon-based decision making theoretically provides the maximum total reward[16]. However, as discussed in the introduction, decision conflicts could yield greater total reward depending on the reward environment conditions. Here, we define the notion of a *happy hour*. During the happy hour, one of the two slot machines dispenses a reward of unity per play to all players who select that machine even when the decisions are conflicted, as schematically shown in Fig. 2b. In the present study, we assume that the higher-reward-probability machine occasionally provides a happy hour. This means that a player gets one coin when he wins even if the decision is in conflict with the decision of the other player during the happy hour. At the same time, it should be emphasized that the reward per play is unity during non-happy hours. That is, a player can get one coin when he wins if the decision is not conflicted. Therefore, the player *cannot* detect the occurrence of a happy hour simply by observing the amount of reward per play. On the other hand, the player can immediately realize the end of the happy hour because the dispensed reward decreases to one-half due to decision conflict.

Mixed strategy. The aim of the present study is to statistically *mix* entangled-photon-based decision-making and correlated-photon-based decision making. While the entangled photons provide non-conflict decisions, the players can accumulate information about the slot machines. Assume that Machine i is selected N_i times, and the number of wins is L_i. Based on maximum likelihood estimation, the estimated reward probability of Machine i is given by $\hat{P}_i = L_i/N_i$ (i = A, B). Here, we consider that the machine that gives the maximum $\hat{P}_i$ would be the best machine. This machine is denoted as Machine m. The source photon states are switched (by $\mathrm{HW_E}$ and $\mathrm{QW_E}$ in Fig. 1a) so that they provide correlated photons. Although here we tune a common photon pair source for both types of states, this could be done by switching from one distinct photon pair source to another, without any difference for the results. At the same time, the half-wave plates of Players 1 and 2 (denoted, respectively, by HW_1 and HW_2 in Fig. 1a) are configured in such a way that Machine m is chosen based on Eqs. (4) and (5), i.e., conflicting decision-making is intentionally induced. If the amount of reward is unity in such an intentionally induced conflicted decision, we can deduce that Machine m is indeed being operated at a happy hour. In addition, once Machine m returns to the non-happy hour operation, the player can immediately detect the end of the happy hour since the dispensed reward becomes one-half due to decision conflict. Such a mixed strategy of entangled- and correlated-photons is summarized in Algorithm 1 and Fig. 2c.

Algorithm 1: Entangled and correlated-photon mixed strategy	
1.	[Entangled strategy] Play slot machines based on entangled-photon-based decision making while accumulating knowledge about the reward probability of the slot machines. Repeat this strategy during *SI* steps. *SI* refers to search interval. Determine the highest reward probability machine by $m = \arg\max_i(\hat{P}_i)$.

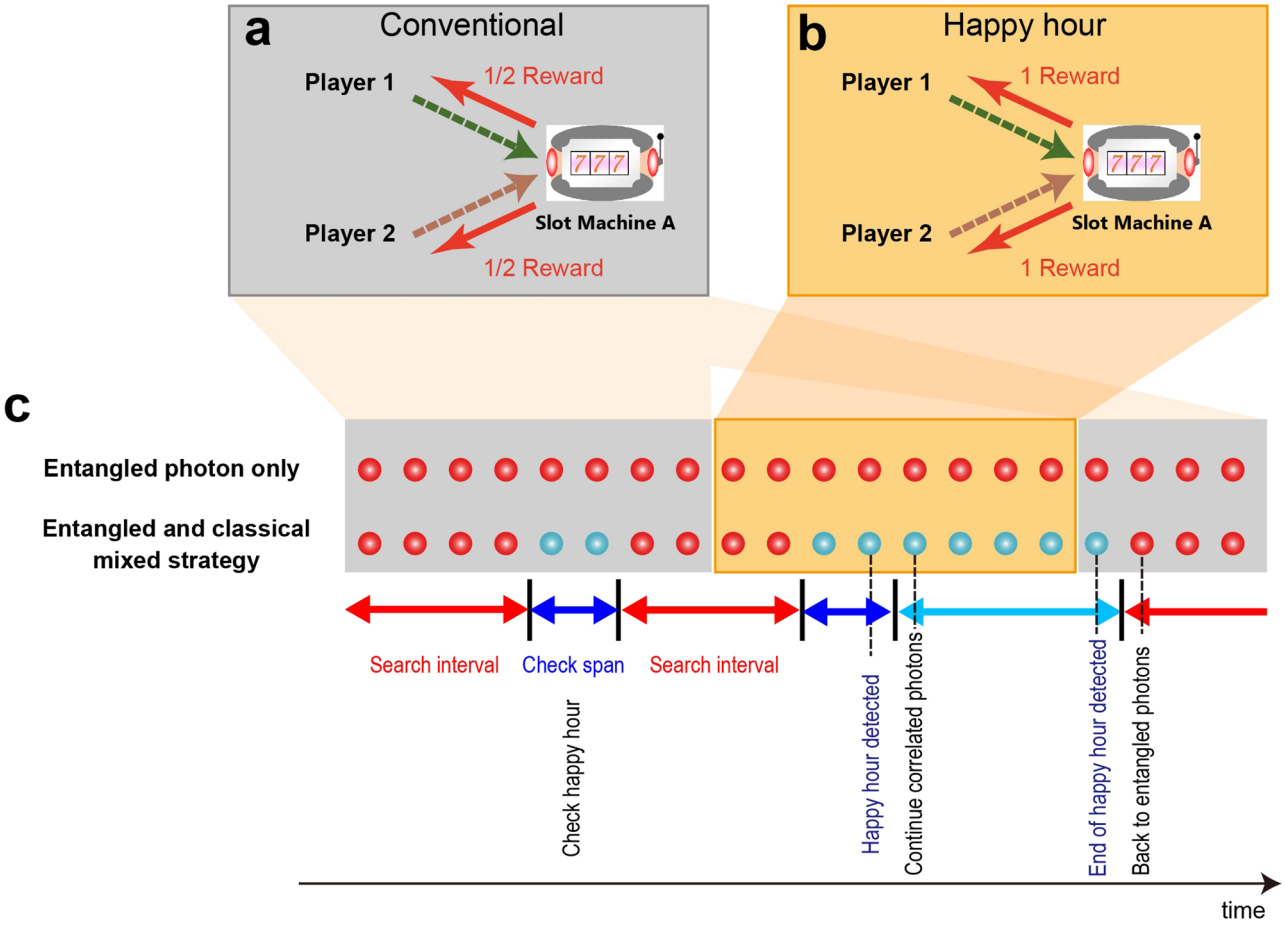

Figure 2. Reward environments and collective decision-making strategies. (**a**) Usually, the reward gained by a single player is half a unit reward, if the decision is conflicted. (**b**) During the happy hour, the higher-reward probability machine yields unit reward to both the players, even if the decision is conflicted. (**c**) In the mixed strategy, while entangled photons are used most of the time, correlated photons are occasionally used (denoted by search interval) during certain duration (check span) to check if greedy action is more beneficial.

Algorithm 1: Entangled and correlated-photon mixed strategy	
2.	[Correlated strategy] Play slot machines based on correlated-photon-based decision-making. Here, the half-wave plates of Players 1 and 2 are configured so that both players select Machine *m*. Repeat this strategy during *CP* steps. *CP* refers to check span.
3.	If the dispensed reward does not become unity, go back to the entangled-photon decision making (Step 1). If the dispensed reward is unity (i.e., Machine *m* is operated in a happy hour), the correlated-photon strategy is maintained. When the dispensed reward becomes half, go back to the entangled-photon decision making (Step 1).

Discussion

The present study is stimulated by the notion of an evolutionary stable strategy (ESS) known in evolutionary game theory[24]. Let us roughly describe the Hawk-Dove game to convey a fundamental concept of ESS. Detailed discussions of this strategy are available in literature, such as Ref.[24]. Players who choose the Hawk strategy always take adversarial actions when confronted by their opponents. By doing so, they can gain a lot of rewards if they win the battle. However, they could also suffer from huge damage if they lose. Conversely, players who choose the Dove strategy always avoid battles when they face their enemies. Hence, there is no gain (because they avoid battles), but there is also no risk of loss. In evolutionary game theory, there exists an optimal mixture of Hawk and Dove strategies that maximizes the expected rewards, and this mixed strategy can be superior to both pure Hawk and pure Dove strategies depending on the environment. The optimal mixture depends on the gains and losses in the battle.

We observe similarities between this concept in evolutionary game theory and the present study of the CMAB problem. The Dove strategy is similar to the entangled-photon strategy which attempts to secure the achievable total reward, while the Hawk strategy is like the correlated-photon strategy which seeks greater reward at a certain degree of risk. The difference lies in the method by which the optimal mixture is derived.

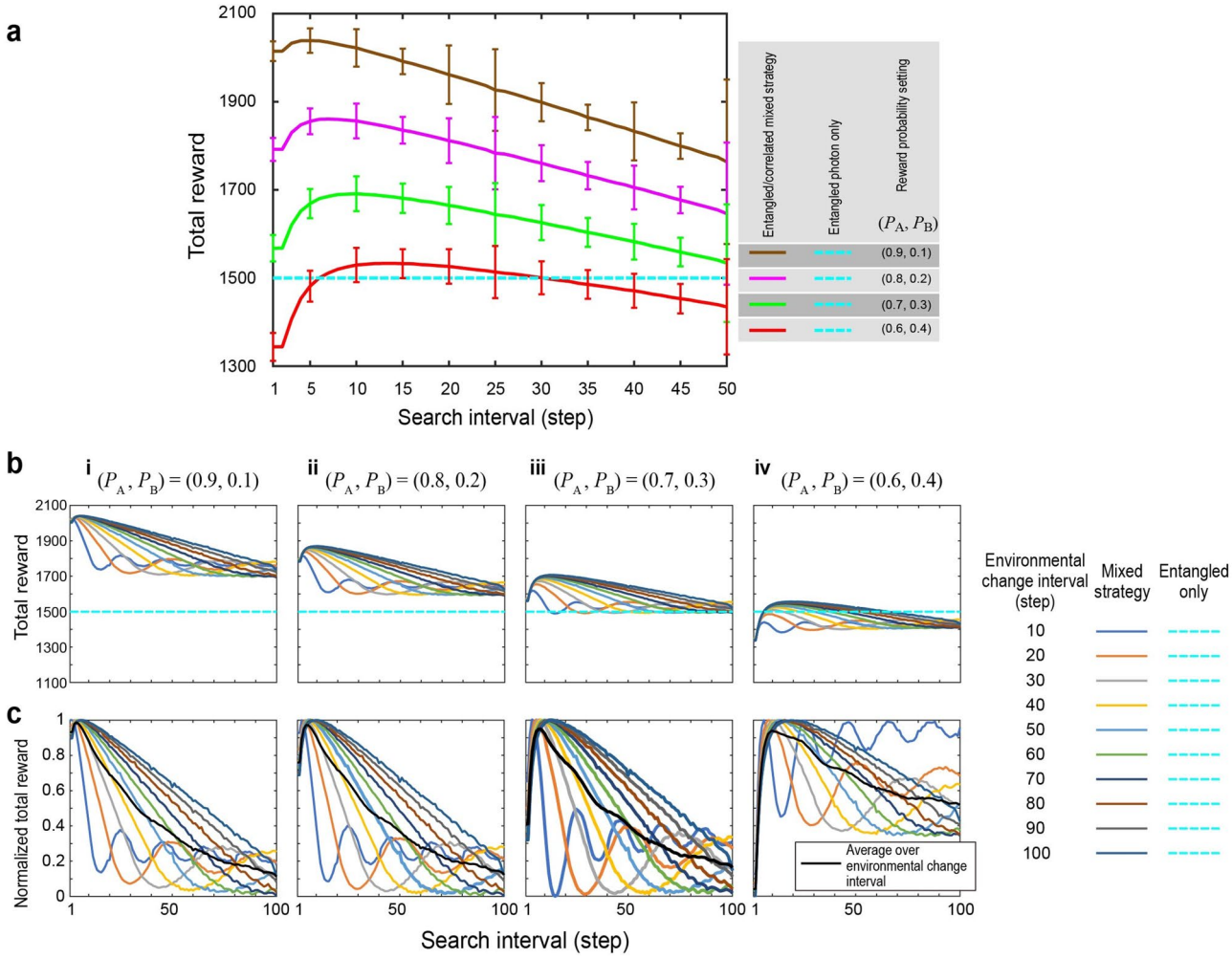

Figure 3. Demonstration of the proposed mixed strategy. (**a**) The dashed cyan line shows the total reward gained by using the entangled-photon-only strategy. With the mixed strategy, the total reward can be greater than that of the entangled-photon-only decision-making strategy. Here, the happy hour and non-happy hour are periodically switched every 50 steps. (**b**) The dependence of the total reward on the environmental change interval, and (**c**) their normalized representation to examine optimality of the mixture of entangled and correlated photons for a variety of environmental conditions.

In the following numerical analysis, 1500 consecutive slot machine plays are conducted for the 2-player, 2-armed CMAB problem. For the sake of simplicity, we assume a fixed reward probability throughout the 1500 plays for P_A and P_B. The total reward, which is the summation of the rewards gained by Player 1 and Player 2, is calculated by averaging over 1000 repetitions of such 1500 consecutive plays. See the "Methods" section for the details. In addition, we assume that P_A is greater than P_B while the condition of $P_A + P_B = 1$ holds. Therefore, if there are no happy hours, the expected maximum total reward is 1500 by the entangled-photon decision strategy. This is because the entangled photons ensure the absence of conflict, meaning that the two machines are always selected. Conversely, the condition of $P_A + P_B = 1$ leads to a constant total reward by entangled photon only strategy, which allows us to examine the effect of the mixed strategy only. The dashed blue line in Fig. 3a shows the calculated total reward.

Now, we examine the impact of the occurrence of happy hours. We assume that happy and non-happy hours periodically interchange in every T steps, with T being an integer. Let us first focus on the case when T is equal to 50. The red curve shown in Fig. 3a represents the average total reward as a function of the search interval when (P_A, P_B) = (0.6, 0.4). We can observe that the total reward is greater than that of the entangled-photon-only strategy if SI is between 6 and 30, while the maximum total reward is realized when $SI = 14$. Hence, $SI = 14$ provides the optimal mixed strategy for this particular reward environment. A search interval that is too short indicates excessive greedy actions ($SI < 6$), whereas an excessively long search interval indicates missing a large reward during the happy hours ($SI > 30$). For the numerical evaluation, the initial starting time of the happy hour was determined for each repetition by a uniformly distributed random natural number between 1 and $2 \times T$, where T is the interval of happy and non-happy hours. The solid curve is the average of the total reward over 1000 such randomly arranged repetitions. The error bars show the corresponding standard deviation, which was found to be unaffected by larger repetition numbers (up to 10,000). Note that the standard deviation is indicated

only for the search intervals of 1 and $5 \times n$ ($n = 1, \ldots, 10$) in Fig. 3a. The computing environments are described in the "Methods" section.

The green, magenta, and brown curves in Fig. 3a show the average total reward when (P_A, P_B) is equal to (0.7, 0.3), (0.8, 0.2), and (0.9, 0.1), respectively. The optimal search interval that yields the maximum total reward decreases as P_A becomes larger. This is because the rewards gained during happy hours dramatically increase as P_A increases. When $P_A = 0.9$, the total reward is almost 2050, which is nearly a 40% increase compared with the entangled-photon-only strategy. In addition, when P_A is greater than 0.7, the total reward is greater than the entangled-photon-only strategy even with extremely small as well as extremely long search intervals, indicating that the gain accomplished during a happy hour surely pays off the cost of conducting correlated-photon-based greedy actions.

The optimal mixture, however, also depends on the frequency of happy hours. The curves in Fig. 3b examine the total rewards when the interval of the happy hour is increased from 10 to 100 steps. First, we consider the case when (P_A, P_B) = (0.6, 0.4) denoted by Fig. 3b-iv. As the happy-hour interval decreases, the maximum total reward decreases, indicating that the mixed strategy cannot adapt to rapid environmental changes. Nevertheless, we also observe that the search interval that yields the maximum total reward decreases as the happy-hour interval decreases, meaning that frequent usage of a correlated-photon-based decision-making strategy provides a greater total reward. The same tendency is observed in different reward probability settings (P_A, P_B) given by (0.9, 0.1), (0.8, 0.2), and (0.7, 0.3), which are summarized in Fig. 3b-i, ii, and iii, respectively. Furthermore, it is interesting to observe the oscillatory behaviour of the total rewards as a function of the search interval. This is clearly due to interdependence between the environmental switching and the strategy switching as the period of oscillation is about twice the period of environmental switching and does not depend on the reward probabilities.

For the sake of obtaining higher rewards regardless of the given environmental change dynamics and to accommodate any uncertainty in the given environment, we discuss the optimality of the search interval. The curves in Fig. 3c represent the normalized total reward, $(R - R_{\mathrm{MIN}})/(R_{\max} - R_{\mathrm{MIN}})$, where R is the total reward for a search interval. R_{MIN} and R_{MAX} indicate the minimum and maximum total rewards in the range of the search interval under study ($1 \leq SI \leq 50$). The black curves represent the average total reward over different happy-hour intervals. The search interval that maximizes the normalized total rewards is summarized in Fig. 4a. The horizontal axis represents the difficulty of finding a higher reward probability machine, which is defined as $1 - (P_A - P_B)$. The reward probability of the combination (P_A, P_B) = (0.9, 0.1) corresponds to a difficulty of 0.2, whereas (P_A, P_B) = (0.5, 0.5) corresponds to a difficulty of unity, which means that the slot machines are identical. One remark here is that Machine A provides happy hour in the case of $P_A = P_B$ while Machine B does not provide happy hour. The optimal search interval monotonically increases as the difficulty of finding a better machine increases. Furthermore, if the difficulty is less than 0.8 (the reward probability differences are greater than 0.2), the search interval of approximately 5–10 is close to the optimal search interval. This is also confirmed by Fig. 3b, suggesting that such a search interval can accommodate the uncertainty of reward environments in terms of both the reward probability values as well as the dynamic change of happy hour occurrences.

The check span is another parameter that influences the optimal mixture of entangled and correlated-photon-based decision-making strategy. Here we focus on the case when reward probabilities are given by (P_A, P_B) = (0.7, 0.3), whose optimal search interval is 9 as observed in Fig. 4a. While keeping this search interval ($SI = 9$), the square marks in Fig. 4b show the average total rewards as a function of CP when the happy hour interchange interval is 50. The number of repetitions is 1000 while the initial starting timing of the happy hour is randomly specified for each repetition as in the above analyses. The maximum total average reward is obtained when CP is 2. This result indicates that too short an CP ($CP < 2$) may miss the detection of a happy hour because of the probabilistic attributes of the slot machine, whereas excessive CP ($CP > 2$) also leads to excessive loss. The analysis in Fig. 3 was conducted based on an CP value of 2. However, it should be noted that the achievable rewards are comparable in view of the standard deviation denoted by the error bars. This indicates that CP is not a dominant parameter; hence the optimality of CP cannot conclusively be affirmed.

In the demonstrations above, the summation of the reward probabilities has been kept unity: $P_A + P_B = 1$. Again, this condition is for the purpose of analyzing the effect of the mixed strategy while keeping the same total reward with the entangled photon only strategy. The proposed strategy does work with other reward environment in general. Figure 5 shows the total reward as a function of the search interval for the cases when P_A is given by 0.9, 0.7, 0.5, and 0.3 while P_B is being kept at 0.2. The other conditions are the same with Fig. 3a; the happy and non-happy hour are periodically switched every 50 steps and the check span is 2. The dashed lines depict the total reward when entangled-photon-only strategy is adopted, which is given by $(P_A + P_B) \times 1500$ where 1500 is the total number of plays. The amount of the total reward obtained by the mixed strategy becomes larger than for the entangled-photon-only strategy as P_A increases because the reward obtained during happy hour increases. On the other hand, when $P_A = 0.3$ and $P_B = 0.2$, the merit of the mixed strategy is negligible or even negative since the difference of the reward probability is small, which is similar to the observation in Figs. 3 and 4.

Before concluding the paper, we put a few remarks on this study. In this work, we focussed on the 2-player, 2-machine CMAB problem to highlight the central concept and principle of the entangled and correlated-photon mixed strategy. The extension of the proposed method to a general case, N-player, M-machine CMAB is an important future study. Indeed, Chauvet et al. have already demonstrated optimal entangled photon states for three, four, and five players for the 2-armed bandit problem[21]. Scalability analysis is also critical from the viewpoint of practical applications such as resource management in information and communication infrastructure[10]. Also, the mixed strategy studied herein allows the players to immediately change the photon states from correlated to entangled photons when they detect the end of happy hour in Step 3 of Algorithm 1. Such controllability or accessibility of the photon source by the player can be generalized through using time delay for example, which is an interesting future topic.

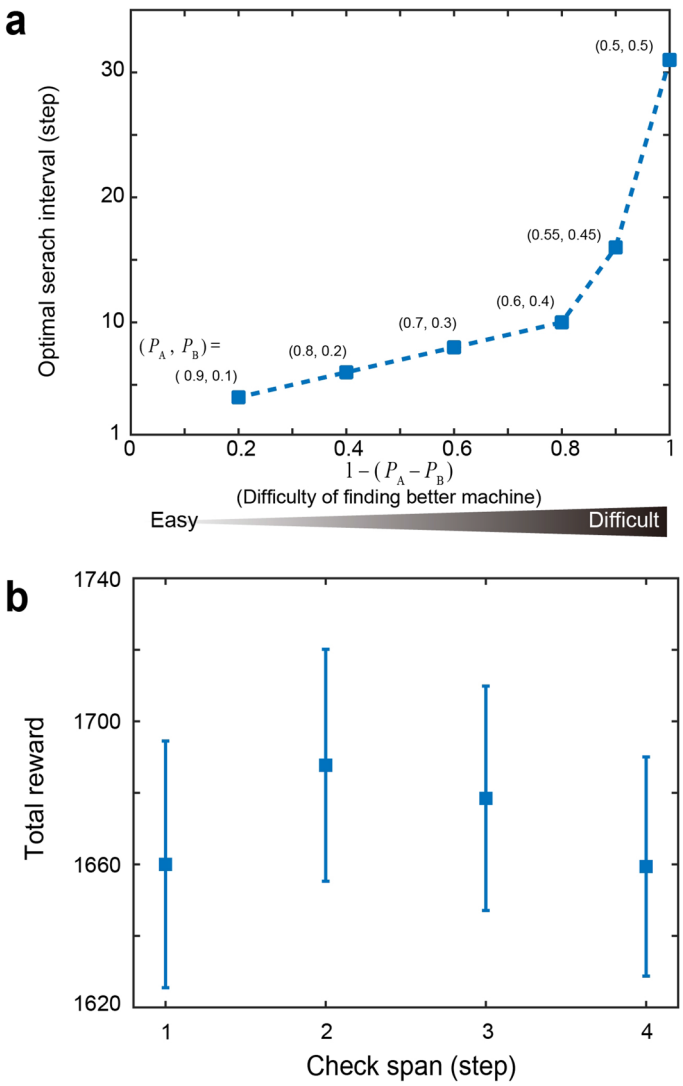

Figure 4. Optimal mixture of entangled- and correlated-photon-based decision making. (**a**) The optimal search interval monotonically increases as the difficulty of finding better slot machine increases. Also, if the difficulty is less than 0.8, a search interval of approximately 5–10 yields nearly optimal total rewards (see Fig. 3b). (**b**) The check span (CP) is another parameter for the mixed strategy. CP of 2 yields the maximum average total rewards. However, concerning the standard deviation denoted by the error bars, the achievable rewards are comparable, indicating that CP is not necessarily a dominant parameter.

The reward probability estimation also needs to be studied. In Step 1 of Algorithm 1, the expected reward probabilities of Machine A and B were evaluated as $\hat{P}_A = L_A/SI$ and $\hat{P}_B = L_B/SI$, respectively, after *SI*-step slot machine plays. Here, L_A and L_B denote the number of wins by playing Machines A and B, respectively. It is remarkable that the denominators of $\hat{P}_A$ and $\hat{P}_B$ are both *SIs* because the decisions were based on entangled photons, i.e., Machine A and Machine B were chosen exactly the same number of times. However, it should also be noted that in the present study, the information integration is assumed from both players. In a future study, we will tackle the case where the reward probability estimation is completely independently conducted by each player. We would also presume, however, that the impact of this independent estimation may be minor, particularly when *SI* is moderately large owing to the equality secured by entangled photons. Meanwhile, an alternative approach to find the higher reward probability machine is to utilize a round-robin scheduling; for instance, Players 1 and 2 select Machines A and B, respectively, during *SI*/2 plays, and vice versa for the subsequent *SI*/2 plays using classical photon pairs. This approach, however, requires a pre-determined coordination among players to avoid conflicting selections, which is outside the problem setting of the present study. Entangled photon pairs, on the other hand, autonomously provide random and non-conflicting choices for the players.

It should also be noted that order recognition is required for solving general *N*-player, *M*-machine CMAB problems, especially when $M > N$. Hence, finding the highest reward probability machine alone is not sufficient, and a novel strategy should be developed for order recognition. In this respect, we presume that the random and non-conflicting selections by entangled photons may provide more efficient recognition of the reward environment, even when compared with the pre-coordinated round-robin scheduling approach. To achieve this, several

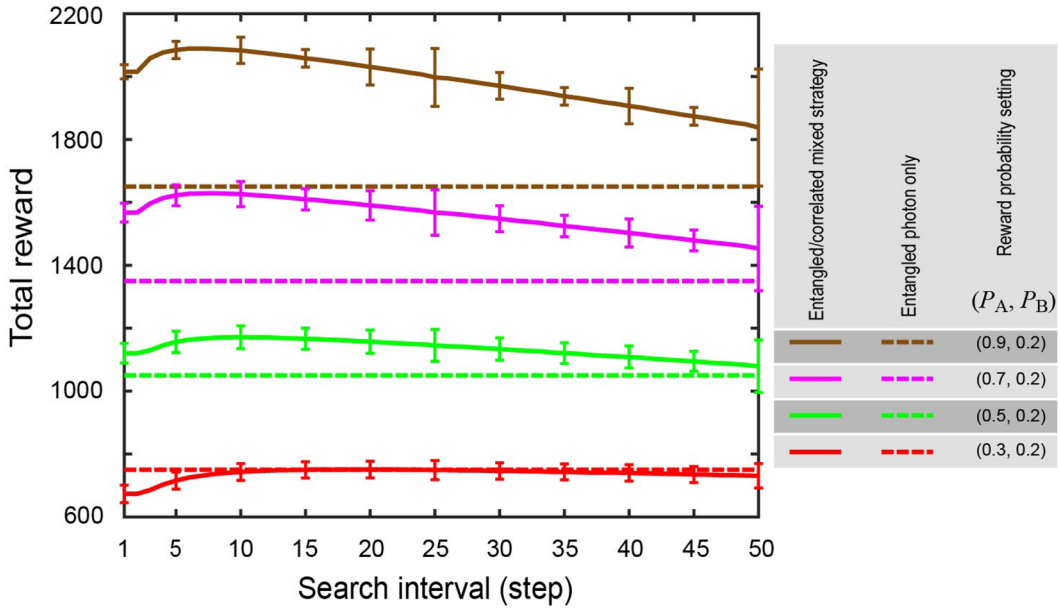

Figure 5. Total rewards by the mixed strategy and the entangled photon only strategy for different general reward environments. The reward environments are given by $P_A = 0.3$, 0.5, 0.7, 0.9 and $P_B = 0.2$. The entangled photon only strategy yields different total rewards depending on the summation of P_A and P_B. The mixed strategy gains greater total rewards than the entangled photon only strategy as P_A becomes larger. On the other hand, the gain is negligible or even negative when the difference between the reward probabilities is small ($P_A = 0.3$, $P_B = 0.2$).

approaches, such as using confidence interval[25] and Schubert calculus[26], could be integrated with the present study in the future.

Conclusion

We theoretically and numerically demonstrated an entangled and correlated-photon-based mixed decision-making strategy to obtain enhanced total rewards in dynamically changing reward environments. Entangled-photon-based decision-making completely avoids conflict and secures equal opportunities for all players. However, conflict avoidance does not necessarily maximize total rewards in reward environments where greedy actions are beneficial, even socially. By introducing the notion of happy hours in competitive multi-armed bandit problems, the cases in which conflicts are beneficial are systematically examined. We demonstrated an optimal mixture of entangled and correlated-photon strategies in terms of adequate switching intervals between these two strategies. The present study is relevant to evolutionary stable states known in evolutionary game theory, where the optimally mixed strategy provides greater expected rewards than other mixed and pure strategies in biological species. We observe similarities between the proposed method and evolutionary game theory in terms of the mixture in strategies themselves as well as the dependence on the given environment. This study paves the way for utilizing both quantum and classical aspects of photons in a mixed manner as well as demonstrating yet again, the supremacy of mixed strategies.

Methods

Numerical analysis. The numerical analysis of the present study was conducted on a personal computer (MacBook Pro, Intel Core i5, 1867 MHz, 8 GB RAM, macOS Catalina, MATLAB R2019a). For emulating the entangled photons and the two slot machines, we utilized uniformly distributed pseudorandom numbers generated by Mersenne twister.

Data availability

Data used in this study is available upon reasonable request to the corresponding author.

Received: 20 October 2020; Accepted: 15 February 2021
Published online: 01 March 2021

References

1. Kitayama, K. *et al.* Novel frontier of photonics for data processing—Photonic accelerator. *APL Photonics* **4**, 090901 (2019).
2. Larger, L. *et al.* Photonic information processing beyond Turing: An optoelectronic implementation of reservoir computing. *Opt. Express* **20**, 3241–3249 (2012).
3. Brunner, D., Soriano, M. C., Mirasso, C. R. & Fischer, I. Parallel photonic information processing at gigabyte per second data rates using transient states. *Nat. Commun.* **4**, 1364 (2013).

4. Sugano, C., Kanno, K. & Uchida, A. Reservoir computing using multiple lasers with feedback on a photonic integrated circuit. *IEEE J. Sel. Top. Quant.* **26**, 1–9 (2019).
5. Shen, Y. *et al.* Deep learning with coherent nanophotonic circuits. *Nat. Photonics* **11**, 441–446 (2017).
6. Ishihara, T., Shinya, A., Inoue, K., Nozaki, K. & Notomi, M. An integrated nanophotonic parallel adder. *ACM J. Emerg. Technol. Comput. Syst.* **14**, 1–20 (2018).
7. De Lima, T. F., Shastri, B. J., Tait, A. N., Nahmias, M. A. & Prucnal, P. R. Progress in neuromorphic photonics. *Nanophotonics* **6**, 577–599 (2017).
8. Nahmias, M. A. *et al.* Photonic multiply-accumulate operations for neural networks. *IEEE J. Sel. Top. Quant.* **26**, 1–18 (2019).
9. Lai, L., El Gamal, H., Jiang, H. & Poor, H. V. Cognitive medium access: Exploration, exploitation, and competition. *IEEE Trans. Mobile Comput.* **10**, 239–253 (2011).
10. Takeuchi, S. *et al.* Dynamic channel selection in wireless communications via a multi-armed bandit algorithm using laser chaos time series. *Sci. Rep.* **10**, 1574 (2020).
11. Kim, S. J., Naruse, M. & Aono, M. Harnessing the computational power of fluids for optimization of collective decision making. *Philosophies* **1**, 245–260 (2016).
12. Naruse, M. *et al.* Single-photon decision maker. *Sci. Rep.* **5**, 13253 (2015).
13. Flamini, F. *et al.* Photonic architecture for reinforcement learning. *New J. Phys.* **22**, 045002 (2020).
14. Naruse, M., Terashima, Y., Uchida, A. & Kim, S. J. Ultrafast photonic reinforcement learning based on laser chaos. *Sci. Rep.* **7**, 8772 (2017).
15. Ma, Y. *et al.* Time-delay signature concealment of chaos and ultrafast decision making in mutually coupled semiconductor lasers with a phase-modulated Sagnac loop. *Opt. Express* **28**, 1665–1678 (2020).
16. Chauvet, N. *et al.* Entangled-photon decision maker. *Sci. Rep.* **9**, 12229 (2019).
17. Piccinotti, D., MacDonald, K. F., Gregory, S., Youngs, I. & Zheludev, N. I. Artificial intelligence for photonics and photonic materials. *Rep. Prog. Phys.* **84**, 012401 (2020).
18. Genty, G. *et al.* Machine learning and applications in ultrafast photonics. *Nat. Photonics* https://doi.org/10.1038/s41566-020-00716-4 (2020).
19. Sutton, R. S. & Barto, A. G. *Introduction to Reinforcement Learning* (MIT Press, Cambridge, 1998).
20. Naruse, M. *et al.* Decision making photonics: Solving bandit problems using photons. *IEEE J. Sel. Top. Quant.* **26**, 7700210 (2019).
21. Chauvet, N. *et al.* Entangled N-photon states for fair and optimal social decision making. *Sci. Rep.* **10**, 20420 (2020).
22. Fedrizzi, A., Herbst, T., Poppe, A., Jennewein, T. & Zeilinger, A. A wavelength-tunable fiber-coupled source of narrowband entangled photons. *Opt. Express* **15**, 15377–15386 (2007).
23. Kok, P. *et al.* Linear optical quantum computing with photonic qubits. *Rev. Mod. Phys.* **79**, 135 (2007).
24. Weibull, J. W. *Evolutionary Game Theory* (MIT Press, Cambridge, 1997).
25. Narisawa, N., Chauvet, N., Hasegawa, M. & Naruse, M. Arm order recognition in multi-armed bandit problem with laser chaos time series. arXiv:2005.13085.
26. Uchiyama, K. *et al.* Generation of Schubert polynomial series via nanometre-scale photoisomerization in photochromic single crystal and double-probe optical near-field measurements. *Sci. Rep.* **10**, 2710 (2020).

Acknowledgements
This work was supported in part by the CREST project (JPMJCR17N2) funded by the Japan Science and Technology Agency, the Core-to-Core Program A. Advanced Research Networks and Grants-in-Aid for Scientific Research (A) (JP20H00233) funded by the Japan Society for the Promotion of Science.

Author contributions
M.N. and N.C. directed the project. S.M., M.N., and N.C. designed the system architecture. S.M. and M.N. designed and built a theoretical fundamental on the mixed strategy system. S.M conducted the numerical analysis. N.C., G.B., S.M., and M.N. examined the physical modelling. N.C., H.S., H.H., G. B, S.H., and M.N. discussed the results. S.M. and M.N. wrote the paper. All authors have revised the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to S.M. or M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2021