Effect of a corrupted source in Samaritan's dilemma

Junichi Shimamura[1] * Şahin Kaya Özdemir[1] †
Fumiaki Morikoshi[2] ‡ Nobuyuki Imoto [1 2] §

[1] *CREST Research Team for Interacting Carrier Electronics, The Graduate University for Advanced Studies (SOKENDAI), Hayama, Kanagawa 240-0193, Japan*
[2] *NTT Basic Research Laboratories, NTT Corporation, 3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan*

Abstract. We investigate the Samaritan's dilemma [1], which is also known as the Welfare Game, within quantum game theory in the presence of a corrupted source. Quantum games have been studied assuming a perfect source which produces a known pure input state with probability one. Under these conditions, it has been shown in several games that the dilemma in the classical versions can be resolved. On the other hand, our studies reveal that the strategies and payoffs in the Nash equilibrium (NE) depend heavily on the initial state.

Keywords: quantum game, corrupted source, Samaritan's dilemma, welfare game

In recent years, there has been a steadily increasing interest in game theory within the quantum information community. It has been shown that the dilemmas existing in several classical games can be resolved by using the quantum mechanical toolbox (quantum operations and entanglement) [3]. The types of dilemmas that have been explored and solved using the paradigm of quantum mechanics can be classified into three groups according to their payoff matrices in pure classical strategies; (i) multiple NE (e.g. Battle of sexes, Stag-Hunt games [2]) (ii) only one NE, which does not coincide with the Pareto optimal (e.g. Prisoner's dilemma [3]), and (iii) absence of NE (e.g. Samaritan's dilemma [4]).

In this study, we discuss the Samaritan's dilemma, whose payoff matrix for classical strategies is given in Table 1 [1]. In economy modeling, the relation between the players of this game corresponds to a situation where the player (Alice or the Samaritan) wishes to aid the person in need (Bob) if he searches for work but not otherwise, and Bob searches for work only if he cannot depend on Alice's aid. It is seen from this table that for the classical pure strategies there is no NE in the game. Under classical mixed strategies, an NE appears with a payoff $(\$_A, \$_B) = (-0.2, 1.5)$ when Alice chooses "*Aid*" with probability 0.5 and Bob chooses "*Work*" with probability 0.2. Although the classical mixed strategy gives a unique NE, Alice's payoff is less than Bob's one and also negative. This is Samaritan's dilemma, which corresponds to the fact that Alice who likes to help voluntarily to the person in need (Bob) is exploited by the selfish behavior of Bob and cannot stop this exploitation. Eisert *et al.* [3] proposed the quantization scheme shown in Fig. 1 for two-player simultaneous move games. In these games, there is no classical communication between the players. We apply this scheme to the welfare game and discuss the effects of a corrupted source. In this physical model, starting from an initial product state $|fg\rangle$ with $\{f, g\} = \{0, 1\}$, the referee prepares the entangled state

	Bob: Work (W)	Bob: Loaf (L)
Alice: Aid (A)	$(3, 2)$	$(-1, 3)$
Alice: No Aid (N)	$(-1, 1)$	$(0, 0)$

Table 1: Payoff matrix for Welfare Game.

$\hat{\rho}_{\rm in} = \hat{J}|fg\rangle\langle fg|\hat{J}^\dagger$ where $\hat{J}$ is the entangling unitary operator with $\hat{J}|fg\rangle = \frac{1}{\sqrt{2}}[\,|fg\rangle + i(-1)^{(f+g)}|(1-f)(1-g)\rangle\,]$. Then the referee sends one of the qubits of $\hat{\rho}_{\rm in}$ to Alice and the other one to Bob. Alice and Bob choose their local operations $\hat{U}_A$ and $\hat{U}_B$ from the SU(2) operator set, and perform them on their own qubits separately. In particular, two classical operations, the identity and the bit flip, are defined as $\hat{\sigma}_0$ and $i\hat{\sigma}_y$ operations, respectively, in this quantization scheme. After these operations the resulting state becomes $\hat{\rho}_{\rm out} = (\hat{U}_A \otimes \hat{U}_B)\hat{\rho}_{\rm in}(\hat{U}_A^\dagger \otimes \hat{U}_B^\dagger)$. The referee who receives this final state first performs $\hat{J}^\dagger\hat{\rho}_{\rm out}\hat{J}$ and then makes a projective measurement represented by $\{\Pi_n = |j\ell\rangle\langle j\ell|\}_{\{j,\ell=0,1\}}$ with $n = 2j + \ell$ corresponding to the projection onto the orthonormal basis $\{|{\rm AW}\rangle, |{\rm AL}\rangle, |{\rm NW}\rangle, |{\rm NL}\rangle\} = \{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$. According to the measurement outcome n, the referee assigns each player payoffs chosen from the payoff matrix of the classical game, i.e., for the Samaritan's dilemma payoff matrix given in Table 1. Then the average payoff of the players can be written as $\$_A = \sum_n a_n {\rm Tr}(\Pi_{\rm n}\hat{J}^\dagger\hat{\rho}_{\rm out}\hat{J}), \$_B = \sum_n b_n {\rm Tr}(\Pi_{\rm n}\hat{J}^\dagger\hat{\rho}_{\rm out}\hat{J})$ with $a_{\{n=0,1,2,3\}} = \{3, -1, -1, 0\}$ and $b_{\{n=0,1,2,3\}} = \{2, 3, 1, 0\}$ being the payoffs of Alice and Bob in Table 1.

In a recent study, we have shown that when the input state is $|fg\rangle = |00\rangle$, there is a unique NE that corresponds to the strategy $(\hat{U}_A, \hat{U}_B) = (i\hat{\sigma}_z, i\hat{\sigma}_z)$ with the payoff $(\$_A, \$_B) = (3, 2)$, thus solving the dilemma in the game [4]. On the other hand, when the referee starts with a state $|fg\rangle = |01\rangle$, two NE points appear with equal payoffs $(\$_A, \$_B) = (3, 2)$ for their actions $(\hat{U}_A, \hat{U}_B) = (\hat{\sigma}_0, i\hat{\sigma}_y)$ and $(\hat{U}_A, \hat{U}_B) = (i\hat{\sigma}_y, i\hat{\sigma}_z)$.

Now let us assume that the source is corrupted and outputs the state $|\,00\rangle$ with probability p and $|\,01\rangle$ with probability $1 - p$. Then how does this corrupted source affect

*shimamura_junichi@soken.ac.jp
†ozdemir@soken.ac.jp
‡fumiaki@will.brl.ntt.co.jp
§imoto@soken.ac.jp

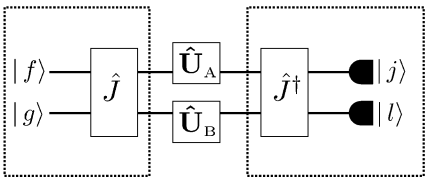

Figure 1: Schematic configuration of the quantization scheme for 2×2 strategic games. The operations inside the dotted boxes are performed by the referee.

the strategies and the payoffs of the players ? We look at the problem in two different situations: (i) The players and the referee know the characteristics of the corrupted source and (ii) they do not know that the source is corrupted and thus believe it is an ideal one.

In the first situation, we look for the best strategies of the players for different values of p and discuss how this corrupted source affects the dynamics of the NE in the game. With this corrupted source, the state distributed to the players by the referee becomes $\hat{\rho}_{\rm in} = p\hat{J}|00\rangle\langle 00|\hat{J}^\dagger + (1-p)\hat{J}|01\rangle\langle 01|\hat{J}^\dagger$. This modifies the payoffs of the players, which in turn, results in the emergence of new multiple NEs for different values of p. The problem becomes very unpleasant for the players because an increase in the number of NEs prevents the players from resolving the dilemma. To give an idea on the strategies of the players which give rise new NEs, we listed some of them in Table 2. It is observed that a unique enforcing NE emerges when $p = 1$, which corresponds to an ideal source that always outputs the state $|00\rangle\langle 00|$. On the other hand, although the case $p = 0$ also corresponds to an ideal source which outputs the state $|01\rangle\langle 01|$, there are two NEs in the case and the dilemma of the players remains. This table clearly shows the input state dependence of the strategies and payoffs of the players in the quantum game. Hence, a corrupted source will change the dynamics of the game.

In the second situation, somehow, the players believe that the source is ideal and uncorrupted, thus fix their strategy to $(\hat{U}_A, \hat{U}_B) = (i\hat{\sigma}_z, i\hat{\sigma}_z)$ to receive $(\$_A, \$_B) = (3, 2)$ because this is the self-enforcing unique NE in the ideal case. However, the real initial state is a classical mixture of $|\,00\rangle$ and $|\,01\rangle$ with probabilities p and $1-p$, respectively. Then, their payoffs under the above action become $\$_A = 4p-1$ and $\$_B = 3-p$. To make a comparison with a purely classical situation, we also calculate the effect when the players choose classical mixed strategies. In this case Alice and Bob applies $\hat{\sigma}_0$ with probability 0.5 and 0.2, and $i\hat{\sigma}_y$ with probabilities 0.5 and 0.8, respectively, because this is the self-enforcing unique NE in classical mixed strategies with an ideal source. The comparison of classical and quantum strategies with dependence of payoffs on p is depicted in Fig.2. In this figure, it is seen that (a) Alice can stop Bob from exploiting her with this quantum strategy when $p > 0.8$, for which $\$_A > \$_B$, (b) classical strategies are more robust to changes in the rate of the corruption of the source.

In summary, our study has revealed that a corrupted source changes not only the amount of payoffs but the dynamics of the game, as well.

	$(\hat{U}_A, \hat{U}_B)$	$(\$_A, \$_B)$
$p = 0$	$(\hat{\sigma}_0,\ i\hat{\sigma}_y)$	$(3,\ 2)$
	$(i\hat{\sigma}_y,\ i\hat{\sigma}_z)$	$(3,\ 2)$
$p = 1/4$	$(\hat{\sigma}_0,\ \hat{\sigma}_0)$	$(0,\ 11/4)$
	$(i\hat{\sigma}_y,\ i\hat{\sigma}_z)$	$(2,\ 9/4)$
$p = 1/2$	$(\hat{\sigma}_0,\ i\hat{\sigma}_0)$	$(1,\ 5/2)$
	$(i\hat{\sigma}_0,\ i\hat{\sigma}_y)$	$(1,\ 5/2)$
	$(i\hat{\sigma}_y,\ i\hat{\sigma}_z)$	$(1,\ 5/2)$
	$(i\hat{\sigma}_z,\ i\hat{\sigma}_z)$	$(1,\ 5/2)$
$p = 3/4$	$(\hat{\sigma}_0,\ i\hat{\sigma}_y)$	$(0,\ 11/4)$
	$(i\hat{\sigma}_z,\ i\hat{\sigma}_z)$	$(2,\ 9/4)$
$p = 1$	$(i\hat{\sigma}_z,\ i\hat{\sigma}_z)$	$(3,\ 2)$

Table 2: Strategies of players at the NE points and their corresponding payoffs when the source is corrupted.

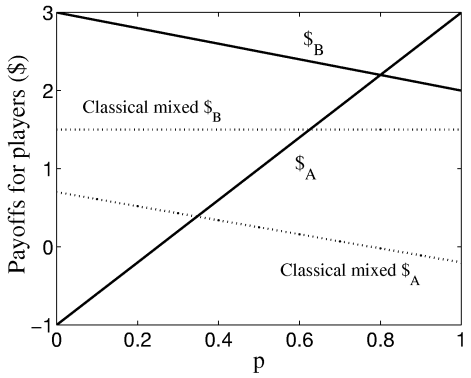

Figure 2: Effect of a corrupted source on the payoffs for Alice and Bob. Solid lines correspond to their payoffs when they apply $(i\hat{\sigma}_z, i\hat{\sigma}_z)$, which gives a unique NE with the initial state $[|00\rangle + i|11\rangle]/\sqrt{2}$, while dashed lines are for the classical mixed strategy.

References

[1] J. M. Buchanan, "The Samaritan's Dilemma," in Edmund Phelps, ed., *Altruism, Morality, and Economic Theory* (New York: Russell Sage), 71 (1975).

[2] J. Shimamura *et al.* in 9th Quantum Information Technology Symposium, Sapporo, Japan, 2003.

[3] J. Eisert, M. Wilkens, and M. Lewenstein, Phys. Rev. Lett. **83**, 3077 (1999).

[4] Ş. K. Özdemir *et al.* in 9th Quantum Information Technology Symposium, Sapporo, Japan, 2003.