Discussion on Some Alternative Interpretations of the Empty Fort Strategy Using Signaling Game and Quantum Cognition

Taige Wang[1]

[1]Department of Mathematics, College of William and Mary, Williamsburg,

10/15/2020

Abstract

As one of the most marvelous strategy in Chinese history, the Empty Fort Strategy allows the extremely disadvantaged defender to survive from the dominant attacker under certain circumstance. The strategist opens the gates and proactively reveals that he or she is capable and willing to pugnaciously resolve the conflict without any conspicuous manpower, which immensely contradicts with the attacker's expectation. The attacker chooses to retreat based on the consideration of an unpredictable ambush inside the fortification. The traditional analysis illustrates that Empty Fort Strategy is a successful example of the application of reverse psychology. We provide an alternative interpretation which categorizes the situation as a two-player signaling game with a pooling equilibrium. We also discuss several minor logic flaws of these two interpretations and how quantum cognition can improve the consistency of the analysis.

1 Introduction

1.1 The Empty Fort Strategy and Examples

Thirty-six Stratagem recorded an example of the Empty Fort Strategy from the *Romance of the Three Kingdoms* [1]. Kingdom Shu and Kingdom Wei were at war. After a great loss of a battle, the prime minister of Shu, Liang was exposed to the commander of Wei, Sima Yi and his army of 150,000 soldiers. With only a few thousand of people, Liang could not win the battle if he chose to attack. However, Liang ordered his soldiers to open the gates and played zither on the platform above the gates with two boys standing behind him. Erratically, Yi ordered a retreat after suspecting that there was a trap inside the city. Although Zhuge Liang was the most renowned exemplar of the strategical thinking, his story about the Empty Fort Strategy was fictitious[2].

However, there are indeed realistic applications of the similarly structured strategy in the history, such as Shu Zhan's strategy recorded *Zuozhuan*[3] and The Battle of Mikatagahara 1573 in Japan[4]. Therefore, it is worth to investigate and analyze this kind of phenomenon. We conclude several characteristics of the Empty Fort Strategy as following:

1. The attacker is significantly more powerful than the defender and sieges the defender's castle and both parties know about this fact.
2. The defender knows the attacker's information. The attacker does not know the specific information about the defender's type, but the attacker believes that the defender is quite disadvantaged.
3. Instead of surrendering or negotiating, the defender chooses to reveal his or her intention to fight the battle that he or she could not possibly win.
4. The attacker suspects that there must have been an unpredictable ambush inside, otherwise it would be illogical for the defender to choose this suicidal option. Therefore, the attacker chooses to retreat to manage the uncertainty.
5. The defender and the attacker are both sophisticated and are able to evaluate the circumstances holistically.

1.2 Traditional Interpretation: Reverse Psychology

One possible reasoning for the success of the Empty Fort Strategy is the reverse psychology, a persuasive technique that the strategist encourages the opponent to act against the strategist's will. Confused by the misleading signals, the opponent makes false judgement about the strategist's true will and chooses to act against the strategist's suggestion, which helps the strategist accomplishes the goal[5].

For instance, a girl chooses to encourage the exact opposite when she needs attention from her boyfriend. Instead of directly asking for it, she would say, "I am fine. I am not mad at you". Usually, some men would choose to stay.

Nevertheless, it leads to a minor logic flaw when we apply it to the Empty Fort Strategy. In the prevalent successful examples of reverse psychology, the strategist is usually the one who is at least as powerful as the opponent or the opponent believes that the strategist is influential enough to cause a potential threat[5]. For our example above, her boyfriend chooses to stay because he believes that there would be a consequence if he chooses to follow the order of the girl. In other words, the opponent believes the consequence of following the order would result in a lower payoff comparing to payoff of the opposite of the order.

In the examples of the Empty Fort Strategy, the attacker generally knows that the defender is rather vulnerable before the defender shows any signals. However, reverse psychology does not explain how revealing strategist's misleading will would change the opponent's belief about the strategist ability to make a credible threat[5]. Therefore, merely revealing the intention to fight does not help to explain that the attacker believes retreating would generate a higher payoff than fight does, because it does not clarify the factor that changes the attacker's belief.

1.3 Perfect Bayesian Equilibrium and Some terms

A perfect Bayesian Equilibrium is a set of strategies and beliefs that are sequentially rational in an incomplete dynamic game. "Dynamic" means players make their decisions sequentially. "Incomplete" means that at least one player is not fully informed about his or her opponent's type, which determines the specific payoff function. "Perfect" means that the threats must be credible. Players update their beliefs whenever it is possible using Bayes' rule[6].

1.3.1 Signaling Game

In a two-person Bayesian game, signaling game is defined when the informed player act first and the uninformed player acts next. The informed player has two possible types that determines the payoffs of both players in different situations. The uninformed player's information is open to the public and thus, the player's type is fixed[6].

1.3.2 Separating Equilibrium

We use the signaling game to model the Empty Fort Strategy and discover a separating equilibrium to interpret it. In a separating equilibrium, which is a type of perfect Bayesian equilibrium, the uninformed player's belief of the informed player's type is so assertive that he or she is able to update informed player's type immediately after the informed player's first move[6].

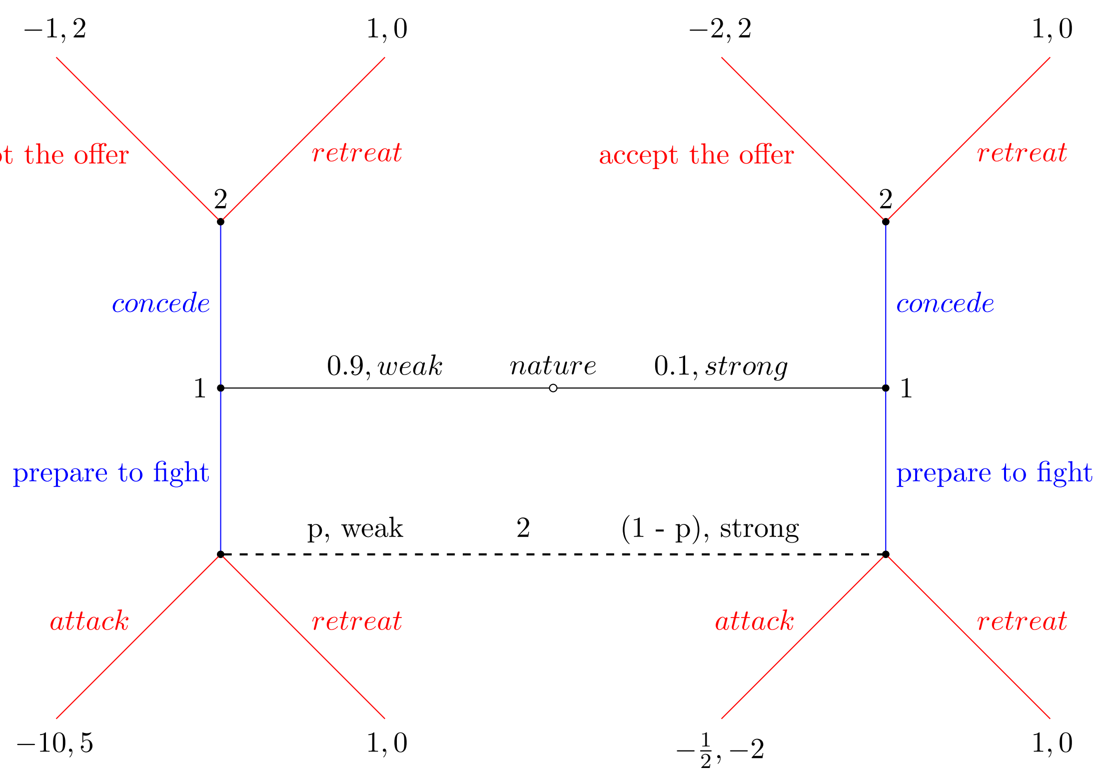
1.3.3 Interim dominance

A strategy is interim dominant strategy if it generates the highest payoff for that specific type of the player regardless of the types of other players[6].

2 Signaling Game Model

2.1 Model Design

We design a two-player signaling game with a separating equilibrium. The informed player (player 1) is the defender who has two possible types (weak or strong) and the player knows all the information about his type, strategy, and the corresponding payoffs of all possible outcomes. The uninformed player (player 2) is the attacker who has a predetermined belief that the defender is a weak type player. The extensive form model looks like as the following:



2.2 Information, Type, Belief, and Strategy

Starting from the nature node, the defender (player 1) acts first. The defender knows that he or she has a high probability of being the weak type and a low probability of being a strong type (unexpected reinforcement). Each type has two strategies, to concede or to reveal the intention that the defender is prepared to fight.

If the defender chooses to concede, the defender's type is exposed to the attacker and the attacker can choose to accept the offer or to retreat. If the defender chooses to reveal that he or she is ready to fight, the attack would not know the defender's type, but the attack can choose to attack or to retreat.

p is the probability that the defender is a weak type that the attacker believes. The attacker would make his or her decision based on the belief. The attacker has a strong belief that the defender is a weak type. Therefore, we let p = 1 in this case.

2.3 Preferences, Payoffs, and Outcomes

Notice that the payoffs we assign here simply display the preference of players. The numbers themselves does not mean anything but how relatively desirable the outcome is for each player[6]. There are eight possible outcomes that represents the combination of each person's strategies:

- When the attacker chooses to retreat, player 2 will have nothing and player 1 could survive. Thus, the payoff would be (1,0) for all four outcomes that player 2 chooses to retreat.

- When player 1 chooses to concede and player 2 chooses to accept the offer, player 2 will have a higher payoff comparing to the payoff of retreating. Weak-type player 1 will have a higher payoff comparing to that of a strong-type player 1 because the weak-type player benefit more from conceding.

- When player 1 chooses to show his or her intention to fight and player 2 chooses to attack, weak-type player 1 will have the lowest payoff among all possible outcomes because player 1 will be dead. Player 2 will have his or her lowest payoff if player 1 is the strong type.

2.4 Separating Equilibrium

Since the attacker (player 2) immediately knows the type of the defender (player 1) if player 1's first move is to concede and accept offer is the interim dominant strategy for player 2, player 2 would choose to accept the offer regardless of player 1's type if player 1 chooses to concede.

Recall that we let p = 1 which means player 2's belief that player 1 is a weak type. Therefore, when a strong-type player 1 chooses to fight, the payoffs are irrelevant to player 2's decision because player 2 would not believe player 1 is a strong type. Player 2 would choose to attack rather than retreating if player 1 chooses to show his or her intention to fight.

Based on that, if player 1 is a weak type, player 1 would choose to concede to have a payoff of -1 instead of -10. Player 1 would also choose to fight to have a payoff of $-\frac{1}{2}$ instead of -2, if player 1 is a strong type.

Hence, the separating equilibrium for this model is (concede, accept the offer) and (prepare to fight, attack).

3 Results and Analysis

3.1 The Interpretation of the Empty Fort Strategy

Since this is a separating equilibrium, in which the uninformed player is able to determine the informed player's type immediately after the informed player's first choice, the uninformed player does not need to use Bayes' rule to update his belief. This model is quite reasonable to the attacker before the attacker see the "empty fort" part.

However, instead of following the separating equilibrium, the defender in the Empty Fort Strategy proactively shows his type by the actions like opening the gates and calmly reveal that the defender is prepared to fight. The only advantage of the defender is the information asymmetry between the attacker and the defender. That telling the uninformed attacker that the defender is the weak type severely threatens the defender's own life, which is counter-intuitive and unreasonable to the attacker. If the attacker tries to rationalize the defender's action based on the separating equilibrium, he or she would changes the belief of the defender's type. We can recalculate the attacker's strategy if the defender chooses to fight with a belief that the defender has a probability of p being the weak type:

$$\begin{aligned} 5p - 2(1-p) &= 0p + 0(1-p) \\ p &= 2/7 \end{aligned} \tag{1}$$

Given our assigned payoffs above, if $p < \frac{2}{7}$, retreating would result in a higher payoff than attacking could. If the attacker changes his or her belief to that the defender has a probability greater than $\frac{2}{7}$ of being the strong type, retreating is a better option.

3.2 Hindrances

Just like the reasoning of reverse psychology, there is another minor logical flaw that perplexes us. Why does not the attacker collect more information to achieve a complete information game? Since the attacker's army is far more powerful than the defender's army, it should be easy to optimize the information. In fact, in the example of Shu Zhan's story, the attacker practically sent spy to collect more information and the attacker still decided to retreat.

Another interesting topic is what the factors that influence the attacker's belief are. These commanders are experienced and brilliant, so we wish to discover the determinant other than the general reasoning like moods or personal preferences.

4 Discussion

4.1 Quantum Cognition

Plenty of inspiring concepts, such as geometry and stochastic processes, obtained their purely mathematical meaning before people discovered their realistic applications. We assume that quantum cognition is one of them. To provide an explanation for the problem of how the attacker changed the belief after the defender showed the signal, we will adopt the mathematical idea from quantum mechanics. There is no physical corresponding evidence that can demonstrate that neurons in our brain are working collaboratively to form quantum states[7]. However, it is quite reasonable to model human cognition based on quantum mechanics.

First, our feelings are constantly changing[7]. For example, when someone asks you how you feel, your feeling may be exhausted when you get up but excited after you brush the teeth. Second, feelings are not always there. To create a feeling, We must intentionally think about it or someone else helps us remind of it[7]. Combining these two points, human cognition is initially at a superposition that multiple states exist at the same time, and it collapses to a state when we observe it. At last, judgement and opinions from other can disturb our own judgement[8]. For example, a man initially prefers to wear a blue tie to participate in the conference. After he knows that his wife prefers the red tie to the blue tie, he might think that red tie looks better.

4.2 System State and Quantum Probability

Relating to cognitive science, we usually define a system to be a person. A state of a system helps to determine the quantum probability of an event that the system generates[9]. In classic probability theory, we essentially use the set theory to model events. The sample space is the set of all of possible elementary events that are mutually exclusive. For example, we can define the set $S = \{\ A, B\ \}$, where A means the defender is the weak type and B means the defender is the strong type. $A \cap B$ represents the probability that the defender is either a weak type or a strong type[10].

Moreover, a state is a probability function that maps the elementary events to the $[0, 1]$. Let E be an event. P(E) $= 0$ represents the event will never happen and P(E) $= 1$ means the event will certainly occur. All of the probability of the elementary event accumulate to 1[10].

In quantum probability theory, we use vector space as the sample space[9]. The event that the defender is a weak type corresponds to basis vector $|A\rangle$. Equivalently, the event that the defender is either a weak or a strong type is the span of $|A\rangle$ and $|B\rangle$.

Furthermore, a state in quantum world is a unit-length vector in N-dimensional vector space. We can change the state by projecting the original state onto the basis vectors that represent the observed events. This is how we perform the process of collapse. The probability of an event given an original unit-length vector S is the squared length of the projection[9]. For example, let

S $= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ and let A $= \begin{bmatrix} 1 & 0 \end{bmatrix}$ be the basis vector

Then, we can changed state from $|S\rangle$ to $|A\rangle$, if the defender is a weak type:

$|A\rangle \langle A| \, |S\rangle = \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{1}{\sqrt{2}} \, |A\rangle$

The probability that the original state collapses to A is $\frac{1}{2}$

4.3 Compatibility and Uncertainty Principle

Compatibility is potentially the most helpful feature to our problem in the quantum cognition model. In classic theory, we do not need to consider the compatibility of events. If event A and B are the elementary events in the sample space, $A \cap B$ must be in the sample space. This means that events are always assumed to be compatible[10].

In quantum theory, the conjunction of event A and B is the intersection of the subspace. However, we could not always simultaneously make judgements about two things. For example, if the attacker could not think about the defender's type and the credibility of the defender's signal at the same time, these two judgements are incompatible. Therefore, they could not share a common basis and we must evaluate them sequentially[8].

Let A = $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and B = $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ form a basis that span R^2.

Suppose there is another basis X = $\begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{\sqrt{2}}{\sqrt{3}} \end{bmatrix}$ and Y = $\begin{bmatrix} \frac{1}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} \end{bmatrix}$

We can describe vector $|X\rangle$ with respect to the $|A\rangle, |B\rangle$ as:

$$|X\rangle = \frac{1}{\sqrt{3}} |A\rangle + \frac{\sqrt{2}}{\sqrt{3}} |B\rangle \tag{2}$$

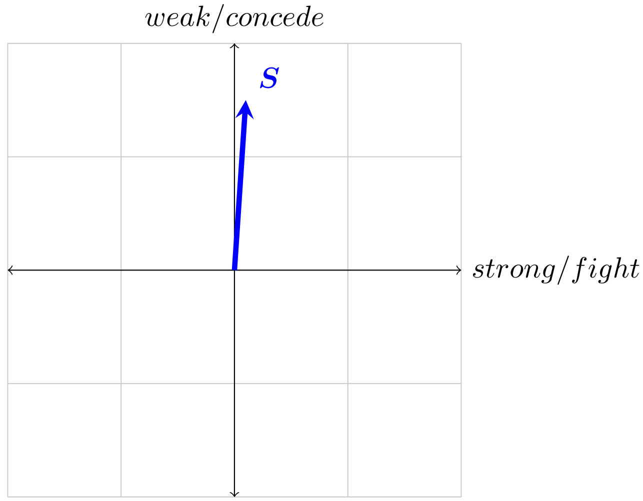
We can also describe vector $|A\rangle$ with respect to the $|X\rangle, |Y\rangle$ as:

$$|A\rangle = \frac{2\sqrt{3} + \sqrt{2}}{2} |X\rangle + \frac{-\sqrt{10}}{2\sqrt{3}} |Y\rangle \tag{3}$$

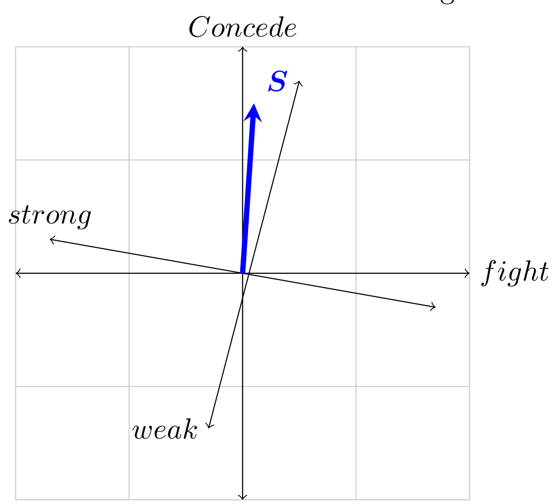
If we are certain about event A, which means that the original state collapses to A, we must be uncertain about X and Y. Therefore, when we evaluate two judgements or feelings that are incompatible, we are able to determine the both of them simultaneously.

4.4 Geometric Interpretation of the attacker's judgements

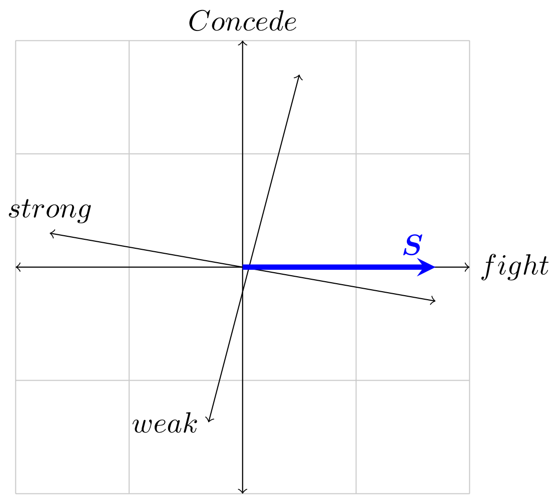
What might be a reason that the attacker changed his or her belief that the defender was a weak type? Initially, based on our separating equilibrium model, the attacker is extremely convinced that the defender is a weak type. Although the attacker did not know the defender's type, but he can easily connected these two judgements about the defender's type and the defender's strategy. In other words, these judgements were compatible previously. Furthermore, the attacker's initial state is close to the weak/concede axis, which represents that there is a quite high probability that the defender would choose to to concede. Therefore, we can graph the state as the following:



However, after the defender proactively shows the intention to fight and they are weak at the same time. The attacker's cognition is disturbed by the additional information. The attacker's belief about the defender's type and the defender's choice is no longer compatible. Thus, we rotate one of the bases to obtain the following:



Since the defender shows the intention to fight, the attacker's cognition must judge defender's decision first. Obviously, the original state collapses to the "fight" basis vector. The revised state looks like this:



Therefore, the revised state has a higher probability of collapsing to the strong state. Hence, the best option for the attacker at the time was to retreat.

In summary, the defender's signal of showing the incompetence and the intention to fight confused the attacker's cognition. The attacker would have to reconstruct the vector space to follow the incompatibility and evaluate the judgements sequentially. Finally, the attacker considers that the probability that the defender of the strong type is high and the attacker chose to retreat.

General

My thanks to Professor Chi-Kwong Li's suggestions.

References

[1] D. Tan. (). "Thirty-six stratagems," [Online]. Available: `https://yabook.org/post/1446.html`. (accessed: 10.15.2020).

[2] G. Luo. (). "Romance of the three kingdoms," [Online]. Available: `https://docs.google.com/file/d/0B6YfeJUvEgsjNGY5MjcxZTgtN2NmNy00ZThhLWIwMWItY2VhMjhmNjRiZTk0/edit`. (accessed: 10.15.2020).

[3] Q. Zuo. (). "Zuo zhuan," [Online]. Available: `https://ctext.org/chun-qiu-zuo-zhuan/zhuang-gong-er-shi-ba-nian/zhs`. (accessed: 10.15.2020).

[4] Kanae. (). "Takeda shingen (35,000) vs tokugawa ieyasu (11,000, including 3,000 oda troops)," [Online]. Available: https://samurai-world.com/the-battle-of-mikatagahara-january-25-1573/. (accessed: 10.15.2020).

[5] G. MacDonald, P. Nail, and J. Harper, "Do people use reverse psychology? an exploration of strategic self-anticonformity," *Social Influence*, vol. 6,2011, pp. 1–14, 2010. DOI: https://doi.org/10.1080/15534510.2010.517282.

[6] N. McCarty and A. Meirowitz, *Political Game Theory*, ser. Analytical methods for social research. Cambridge University Press, 2007, ISBN: 9781107438637.

[7] J. Busemeyer and P. Bruza, *Quantum Models of Cognition and Decision*. Cambridge University Press, 2012, ISBN: 9781107419889.

[8] J. Marciak-Kozlowska and M. Kozlowski, "Heisenberg's uncertainty principle and human brain," *NeuroQuantology*, vol. 11, pp. 1–10, 2013. DOI: 10.14704/nq.2013.11.1.612.

[9] V. F. Jones, *Von Neumann Algebras*. NSF Grant, 2009.

[10] D. Rajan, *Probability, Random Variables, and Stochastic Processes*, ser. Mathematical Foundations for SIGNAL PROCESSING, COMMUNICATIONS, AND NETWORKING. CRC Press, 2012, ISBN: 9781439855133.