

# Theory-of-Mind reasoning in Ancient China

Peter van Emde Boas

*ILLC, FNWI, Universiteit van Amsterdam, PO Box 94242, 1090 GE Amsterdam  
Bronstee.com Software & Services B.V., Heemstede  
peter@bronstee.com*

**Abstract.** Ancient Chinese literature on strategic theory goes back to ideas proposed almost 3000 years ago, but which were first written down during the late Spring and Autumn period and the Warring States period (around 500 BC). The most famous work is the Art of War from Sun Tzu.

While Sun Tzu extensively recommends to prepare before going to war and to study and exploit the strengths and weaknesses of your opponent, he doesn't explicitly mention to investigate the possible reasoning of your opponent. There is no visible trace of Theory-of-Mind reasoning in his work.

More recent literature, like the famous classic historic novel Three Kingdoms (believed to have been written during the 14th Century), which describes events during the period 168 to 280 AD, contains instances of Theory-of-Mind reasoning. The question thus arises when the ancient Chinese did invent Theory-of-Mind reasoning as a concept, and why they seem not to have used it in their extensive Strategic theories.

## Strategic theory in Ancient China, and Sun Tzu

Modern Game Theory, now existing for some 100 years, describes agents in strategic interaction. Strategic interaction in the context of warfare has been studied starting 3000 years ago in China [5], culminating in the famous books by Sun Tzu and Sun Bin [9].

The work by Sun Tzu has an abstract character. The other ancient sources provide more concrete practical advice, aimed at tactics rather than strategy. Noticeable in Sun Tzu's work are the logical features like a semi-structured text and ideas related to modern game theory (*prepare for your enemy, and you may win a war without serious fighting*).

The question arises how close the ancient Chinese came to inventing the fundamental concepts of today's Game Theory. In previous work, Niou and Ordeshook [4] have identified several game theoretical concepts in Sun Tzu. But did the Chinese recognize the inherently circular structure of reasoning about reasoning by opponents? Phrased otherwise: *was Theory-of-Mind reasoning an available tool in Chinese strategic theory building?*

Sun Tzu evidently recognizes the importance of knowing your opponent, as illustrated by the famous quote: *So it is said that if you know others and know yourself, you will not be imperiled in a hundred battles; if you do not know others*

*but know yourself, you win one and lose one; if you do not know others and do not know yourself, you will be imperiled in every single battle.*

It is interesting to observe that Sun Tzu here omits the case specifically recommended by our Western antique knowledge: *Nosce Hostem; Know thy Enemy*; this phrase doesn't involve any knowledge about yourself.

Historical sources give both strategic guidelines, and examples of instances where those guidelines were either used or violated [6, 7]. They contain examples of stratagems - tricks used during warfare to fool your enemy. However, authors of strategy guides, including Sun Tzu, do not explicitly consider the possibility that the enemy may use these strategic recommendations against you. There is no recommendation in Sun Tzu to use Theory-of-Mind reasoning with respect to your opponents. His rules tell you to respond to the opponent for what he is and how he behaves.

### **Theory-of-Mind reasoning**

Generals using higher order Theory-of-Mind reasoning, which is a requirement for seriously taking your opponent's strategy into consideration, do appear in later sources. In the Three Kingdoms novel [8] we meet generals who recognize their opponents' stratagems. This novel describes events dated at the end of the Han Dynasty (168-280 AD), but the book is ascribed to the author Luo Guanzong who lived during the late Yuan and Ming dynasties (the 14th century). The earliest printed edition dates from around 1500 AD. However some of the more interesting events in the book are believed to be fictitious rather than historical.

So the main question is to determine when the ancient Chinese started to use Theory-of-Mind reasoning, specifically in the context of strategic analysis. It is also relevant to specify what we mean by the word "use" here. Johan van Benthem [1] proposes for clarification of this concept (in the context of strategic thinking in general) a three level distinction: *Do it; Understand it; Reason about it*. The difference between the second and third level is that the second level involves pre-theoretical but still reflective reasoning, whereas the third level requires theory building. Evidently, if we are looking for written evidence, this will require at least second level use, since otherwise the authors couldn't have written about it.

A well known instance of Theory-of-Mind reasoning in ancient Chinese philosophy is the encounter between the philosophers Zhuangzi and Huizi:

*One day Zhuangzi and Huizi are strolling on Bridge Hao.*

*Zhuangzi: "Look how happy the fish are just swimming around in the river."*

*Huizi: "How do you know they are happy? You are not a fish."*

*Zhuangzi: "And you are not me. How do you know I don't know the fish are happy?"*

This event is dated during the Warring states period (403-211 AD), which means it is at least a century later than Sun Tzu who lived around 500 BC. But is is much earlier than the events related to the Civil War at the end of the Han Dynasty described in the Romance of the Three Kingdoms.

## The story of the Ambush at Huarong

The most characteristic event in relation to Theory-of-Mind reasoning in the Romance of the Three Kingdoms describes the ambush staged at Huarong on the army of Cao Cao during his retreat after the lost battle at Red Cliffs. The ambush is staged by Guan Yu (Lord Guan) who implements a strategic plan designed by Zhuge Liang (Kongming).

During his retreat Cao Cao indicates two places which would be perfect for an ambush, and indeed on both places Zhuge Liang has prepared one, resulting in severe losses at the side of Cao Cao. Arriving at a junction between roads Cao Cao has two options for continuing his retreat: he can follow a road through the valley, or take a longer more tortuous mountain trail. The scouts he has sent out to recognize these options report that they observe smoke along the mountain trail, indicating the presence of the enemy.

Cao Cao argues that his opponent Zhuge Liang, a well known and intelligent strategist, has organised this sighting of smoke in order to lure Cao Cao into the valley, where he will be ambushed. Strictly in accordance with the teachings of Sun Tzu: *appear weak where you are strong and strong where you are weak*. However, he overlooks the fact that Zhuge Liang has predicted that Cao Cao will argue in this way (also pointing out that the reasoning by Cao Cao is in accordance with the teachings of Sun Tzu).

The result is that Cao Cao takes the mountain road and subsequently is ambushed by Guan Yu. Cao Cao survives by grace of the fact that Guan Yu gives him free passage, in accordance with obligations related to a previous encounter between him and Cao Cao. This is an instance of the warrior's code being observed. According to Mao Zonggang, a 17th Century editor and commentator of the three Kingdoms novel, this escape has also been foreseen by Zhuge Liang, notwithstanding that Guan Yu has promised his head in case he would let Cao Cao escape. But the overall strategic situation makes it advantageous to have Cao Cao not being totally eliminated, so the plan also works perfectly with respect to this higher order issue.

The story evidently exhibits instances of Theory-of-Mind reasoning: Cao Cao uses second order Theory-of-Mind reasoning (as he has done as well on several earlier occasions, for example when he indicates the precise locations for the earlier two ambushes), thus showing that his strategic reasoning goes beyond the basic rules of Sun Tzu. However, Zhuge Liang performs third order Theory-of-Mind reasoning, and thus outsmarts Cao Cao and wins the game. Also the references to the teachings of Sun Tzu are explicit.

There is however a troublesome aspect of this story: it is listed among the fictitious stories, and therefore it can't be used for proving that these instances of Theory-of-Mind reasoning actually did occur in 208 AD. The battle at Red Cliffs and the defeat of Cao Cao are historic facts, but the events during his retreat are not. So the concepts mentioned in the story may have been conceived any time between the actual events and the recording of the Romance of the Three Kingdoms in the 14th century.

## Was Theory-of-Mind reasoning used?

An event believed to be historic which suggests Theory-of-Mind reasoning in the context of Strategic Analysis is the use of the Empty City strategy by Zhuge Liang in 228 AD. Zhuge Liang, having lost a substantial part of his army, hides in the city Xicheng which he possibly can't defend against a stronger enemy. Rather than attempting such a futile defence, he has the gates of the city opened, no troops in sight, and stages a party on top of the city walls, playing his zither in plain sight of the approaching enemy. The enemy, convinced that these open gates are a ruse to lure him into a hidden ambush inside the city, withdraws his troops, and Zhuge Liang escapes.

An game-theoretical analysis of this event can be found in Cotton and Liu [2]. These authors also analyse an earlier similar event dated 144 BC, involving a group of 100 Chinese horsemen against a much stronger army of Barbarian nomads during the Western Han Dynasty. There are several more examples of this strategy mentioned in [3]. Though the event is believed to be historic the description in the Three Kingdoms novel includes fictitious elements; for example the opponent mentioned there, Sima Yi, was not active in that area in the year 228 AD. The Empty City strategy is included as strategy 32 in the traditional collection of 36 strategies of Ancient China [6, 7].

Concluding, it seems likely that the Ancient Chinese, by 200 AD, had discovered the concept of higher order Theory-of-Mind reasoning; however, they didn't have a second abstract thinker like Sun Tzu to discuss it as an independent concept. So there is no evidence for the Chinese having reached the third level of use of Theory-of-Mind reasoning in ancient time. Aside from our mathematical tools which they didn't have, they did not invoke this fundamental ingredient of game theory in their strategic studies. However, they came quite close.

As observed by the reviewers for this workshop, the question on the appearance of Theory-of-Mind reasoning can be raised for other cultures and historical periods as well. I totally agree; however, that would require a far more extended study, which goes beyond the context of the program I am presently involved in: the preparation of a chapter on strategic reasoning for the forthcoming handbook on the history of Logic in China.

## References

1. Johan F.K.A.K. van Benthem, oral. comm. at the 3rd Conference on the history of Logic in China, Tianjin, April 2014.
2. Christopher Cotton & Chang Liu, *100 Horsemen and the Empty City: A Game Theoretic Examination of Deception in Chinese Military Legend*, J. of Peace Research, 48 (2011) 217-233
3. Wikipedia, *Empty Fort Strategy*, [http://en.wikipedia.org/wiki/Empty\\_Fort\\_Strategy](http://en.wikipedia.org/wiki/Empty_Fort_Strategy)
4. Emerson M.S. Niou & Peter C. Ordeshook, *A Game-Theoretic interpretation of Sun Tzu' The Art of War*, J. of Peace Research, 31 (1994), 161-174
5. *The seven Military Classics of Ancient China*, Ed. and transl. Ralph D. Sawyer & Mei-Chün Sawyer, Basic books, NY, New Yourk, USA, 2007

6. Stephan H. Verstappen, *the thirty-six Strategies of Ancient China*, China Books, San Francisco, USA, 1999
7. *The Wiles of War; 36 Military Strategies from Ancient China*, Foreign Language Press, Beijing, 2009
8. Luo Guanzhong, *Three Kingdoms*, vol, I, II and III, ed. and transl. Moss Roberts, Foreign Languages press, Univ. of California press, Beijing, China, 1994
9. Sunzi, *The art of War*, & Sun Bin, *The art of War*, Wu Rusong & Wu Xianlin and transl. Lin Wusun, Library of Chinese classics, Foreign Language press, Hunan peoples publishing house, Beijing, China, 1999