

# Rationality through Threat

## To Catch Hawk, Dove and Chicken <sup>1</sup>

SHIMPEI ENDO <sup>2</sup>

April 7, 2017 (Sat) 15:05-15:55

Are we rational? *We do not (have to) know.*

A THREAT APPEARS ANYWHERE. Look at a parent educating her child to behave well or a warning on legal risk of piracy. Some may further insist that threat is a fundamental mechanism of rule-following and interaction in general.

EPISTEMIC THREAT WORKS SOMETIMES. Roughly, a threat is understood as an assertion of harm <sup>3</sup> which makes <sup>4</sup> others to make certain actions among other options for the sake of utility of who assert it. A direct variant is one declaring (physical, mental, economical or any kind of) harms in a direct manner (e.g. “Treat or (otherwise I will) trick”). This talk will focus on a more indirect type of threat. This version appeals *irrationality* of the speaker herself: “Step back, I am insane”.

INCREDIBLE THREAT. Classical examples (Hawk-Dove and Chicken <sup>5</sup>) contain *threats*. Game theorists have traditionally called it an *incredible threat* and excluded such a move based on the following epistemic assumptions on *rationality*.

EPISTEMIC ASSUMPTIONS ON RATIONALITY. [R1]. Every player is *rational* i.e., every player tries to maximize her own expected utility. [R2]. [R1] above is *common knowledge* among players. <sup>6</sup>

EPISTEMIC THREAT MAY WORK IN A BAD WAY... DUE TO BLUFFING. Does epistemic threat work all the time? No, epistemic threat cannot be the dominant nor winning strategy because of the suspect on *bluffing*. The listener would even take it as an evidence for her rationality the very fact that she commit to such a maximizing option. This worry holds in the other direction. The speaker also worries about the listener’s rationality: whether the target acts as the speaker intends.

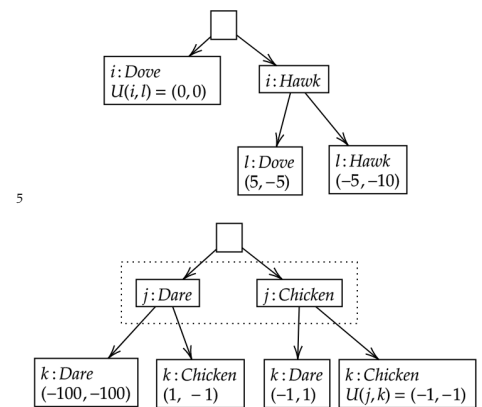
FORMALIZATION VIA DYNAMIC EPISTEMIC LOGIC. There have appeared several approaches toward this idealization of rationality. <sup>7</sup> My framework is to adopt *dynamic epistemic logic*. In DEL, situations of a game are understood as epistemic information and its structure

<sup>1</sup> An individual presentation for the 10th Annual Meeting of Japan Association for Contemporary and Applied Philosophy (JACAP), Nagoya University. The latest version of this handout is available on: <https://www.overleaf.com/read/wrjmhxvswkfx>

<sup>2</sup> Master of Logic, ILLC, University of Amsterdam. My main interests are metaphysics and logic(s), particularly, modal realism and semantics for non-classical modal logics. [endo@shimpeiendo@gmail.com](mailto:endo@shimpeiendo@gmail.com)

<sup>3</sup> Of any kind: physical, mental, economical, and so on.

<sup>4</sup> Rather, it *encourages, leads* or *forces* in a certain context and a pragmatics of the verb.



<sup>6</sup> Common knowledge of R1 is an (infinite) iteration of the follows:  $R_{2_0}$ : everyone knows that [R1],  $R_{2_1}$ : everyone knows that everyone knows that [R1], ... ,  $R_{2_2}$ : everyone knows that everyone knows that everyone knows that [R1]... , and so on.

<sup>7</sup> The most classical attitude ascribes it to our failure to capture the true utility combining any factor (not only economical and physical but also honorary and rumor). Some economists including behavioral ones have suggested new modified solution concepts (somehow, to save the principles of rationality). Other question the rationality principles by importing *type* in the systems to depict the iteration of knowledge (on rationality of each player).

AndréAs. Perea. *Epistemic game theory : reasoning and choice*. Cambridge University Press.

and described in possible world semantics. *Dynamism* is imposed by generating new models or modifying models<sup>8</sup>. You can model simply the *backward induction* which selects an option  $P$  in terms of  $\uparrow P$  (technically, *public announcement*, eliminating any non- $P$  state). We can think of epistemic threat as a more moderate update ( $\uparrow P$ ) on models, which changes probability to each option.

IRRATIONALITY OPERATOR IR. My modification<sup>9</sup> (only) imposes a single operator  $IR_i$ , which corresponds to how *irrational* a player  $i$  becomes.  $IR_i$  changes a probability assigned to each strategy to make. The formal twists are: (i)  $IR_i$  can be *negative* and *positive* and (ii) the actual value of  $IR_i$  stay *unknown* to anybody, including us – who models the situation.

SOLUTION CONCEPT (SKETCH). Even its exact value is unknown (even for the theorist) to anyone, we (theorists, players) can predict and make decision based on given information in *some* cases. For instance, there is no point of considering the possibility of bluffing of the strong (i.e. who takes the initiative in a game, like predator in the Hawk-Dove).

CATCH-22.<sup>10</sup>

(...) There was only one catch and that was Catch-22, which specified that a concern for one's own safety in the face of dangers that were real and immediate was the process of a rational mind. Orr was crazy and could be grounded. All he had to do was ask; and as soon as he did, he would no longer be crazy and would have to fly more missions. Orr would be crazy to fly more missions and sane if he didn't, but if he was sane, he had to fly them. If he flew them, he was crazy and didn't have to; but if he didn't want to, he was sane and had to. Yossarian was moved very deeply by the absolute simplicity of this clause of Catch-22 and let out a respectful whistle. "That's some catch, that Catch-22," he observed. "It's the best there is," Doc Daneeka agreed. (...) (p. 46)<sup>11 12</sup>

## References

- [1] Joseph Heller. *Catch-22, a novel*. The Modern Library, New York, 1961.
- [2] Andrei A. Perea. *Epistemic game theory : reasoning and choice*. Cambridge University Press, 2012.
- [3] Johan van Benthem. *Logic in Games*. The MIT Press, Cambridge, Massachusetts, 2014.

<sup>8</sup> Johan van Benthem. *Logic in Games*. The MIT Press, Cambridge, Massachusetts, 2014

<sup>9</sup> The approach I will *not* take is as follows: to impose a new operator determining *how believable* to a player the threat is.

<sup>10</sup> What (originally) motivated this work include: Cohen Brothers' series for kidnapping, Koji Wakamatsu and the 60-70s new left terrorists which taught the suicide strategy to the world.

<sup>11</sup> What does our analysis with IR operator tell about this scenario? The evaluation of Doc Daneeka may come from the rationality principle which determines who is rational by who successfully picks the most rational (i.e. maximizing her utility) choice.

<sup>12</sup> Joseph Heller. *Catch-22, a novel*. The Modern Library, New York, 1961