Metagame Strategies of Nation-States, with Application to Cross-Strait Relations

Alex Chavez[1] and Jun Zhang[2,*]

[1]achavez@umich.edu, Department of Psychology, University of Michigan, USA
[2]junz@umich.edu, Department of Psychology, University of Michigan, USA

Abstract Metagames (Howard, 1968), a class of formal models in game theory, model the strategic reasoning among players who mutually predict each others actions recursively ("I think you think I think ..."). We present a framework for three-player games based on metagame theory. This framework is well-suited to the analysis of nation-states, especially when the analyst wishes to make few assumptions about the level of recursive reasoning and the preference orderings of players.

1 Introduction

Much of game theoretic analysis relies on the concept of the Nash equilibrium, an outcome from which no player can deviate unilaterally and increase her or his payoff. However, the Nash equilibrium often appears to model stability inadequately in real world applications.

A well-known example in which the solution concept of the Nash equilibrium is inconsistent with behavior is the Prisoner's Dilemma, pictured in Table 1. Although the unique equilibrium in this game is for both players to defect, a typical behavioral finding is that both players will cooperate, even under conditions of one-shot play and strict anonymity.

Table 1 Prisoner's Dilemma. The number of the left in each cell (outcome) is Player 1's payoff, and the number on the right is Player 2's payoff.

		Player 2 Cooperate	Player 2 Defect
Player 1	Cooperate	3,3	0,5
	Defect	5,0	1,1

One recent approach to modeling such discrepancies is to systematically modify the payoff function to include other-regarding preferences such as fairness (Fehr and Schmidt, 1998) or conformity to social norms (Bicchieri, 2006). These approaches have enjoyed considerable success in modeling behavior in the laboratory setting. An alternative better suited for political analysis, however, is to modify the players' strategies instead of their payoffs. Howard (1968)'s theory of metagames (see

* Currently on IPA assignment to AFOSR from the University of Michigan.

also Thomas, 1986) accomplishes this by assuming that one or more players use foresighted, conditional metagame strategies. Using the Prisoner's Dilemma as an example, Player 1's metagame strategies are of the form, "Cooperate if Player 2 cooperates; but defect if Player 2 defects." We refer to these as level-1 strategies. Furthermore, Player 2's strategy set can build off of Player 1's, and include strategies such as "If Player 1 uses the above conditional strategy, cooperate; if player 1 uses a different conditional strategy, defect." We refer to these as level-2 strategies. By specifying the number of nested levels of strategies, and the order in which players form their strategies, one generates the *metagame* corresponding to any normal-form *base game*.

Howard's proposed solution concept, the metaequilibrium, is an outcome in the original game that corresponds to a Nash equilibrium in the metagame. In the Prisoner's Dilemma, it can be shown that there are pairs of level-1 and level-2 strategies that a) are Nash equilibria in the metagame, and b) correspond to (cooperate, cooperate) in the original game, thereby showing that (cooperate, cooperate) is a metaequilibrium. An appealing property of metagames is that including conditional strategies higher than level-n (where n is the number of players) does not change the set of metaequilibria. Thus, there is no need to examine games with conditional strategies beyond level-n.

The metagame approach is useful especially when 1) the modeler is uncertain about the level of conditional strategies being employed by various players, or 2) the players are known to be using higher-level conditional strategies that are not captured in the base game. Both of these conditions often apply to situations of multinational conflict, and Taiwan independence from mainland China is no exception. For this reason, combined with the fact that the status quo in the situation of Taiwan independence is not a Nash equilibrium (as we will show in the following section using minimal assumptions), we use the metagame framework.

All metagames are constructed from a normal form base game. Therefore, we begin in Section 2.1 by defining the situation of Taiwan independence according to the following three required components: the set of *players*, their *strategies*, and their *preferences over outcomes*. In Section 2.2, we formally present the solution concept of the Nash equilibrium, and show that the status quo is not contained in the set of Nash equilibria. Section 3 defines metagame strategies and the metaequilibrium. Then, Section 4 analyzes in detail a particular preference combination in the situation of Taiwan independence, which we have selected for illustrative purposes. Finally, Section 5 concludes with a discussion of robustness of metaequilibria across many different possible preference orderings.

2 The Base Game

2.1 Outcomes and Preferences

We model the situation of Taiwan independence from mainland China as the three-player normal-form game, Γ, in which: China (c), can wage war (*War*) or not (*NoWar*); Taiwan (t) can declare independence (*Independence*) or not (*NoIndependence*); and the United States (u) can support Taiwan (*Support*) or remain neutral (*NoSupport*). These strategies jointly define the set of $2 \times 2 \times 2 = 8$ *outcomes* (see Figure 1).

Each player *i*'s preferences over these outcomes are modeled by a *payoff function*, π_i, which assigns real numbers to the outcomes in a manner that preserves the ordering of preferences. Thus, if Taiwan prefers outcome B to A, then $\pi_t(B) \geq \pi_t(A)$.

U.S.: *Support*

	Taiwan: *NoInd.*	Taiwan: *Ind.*
China *NoWar*	A	B
War	C	D

U.S.: *NoSupport*

	Taiwan: *NoInd.*	Taiwan: *Ind.*
NoWar	E	F
War	G	H

Key
(A) Status quo.
(B) U.S. – recognized independent Taiwan.
(C) Unification with an antagonized U.S.
(D) All – out war.
(E) Status quo with an isolated Taiwan.
(F) U.S. – unrecognized independent Taiwan.
(G) Unification without resistance.
(H) Unification with resistance.

Fig. 1 Taiwan independence as a Three-player Game.

Typically, it is assumed that China prefers outcome G, unification without Taiwan resistance or U.S. support, to all other outcomes. On the other hand, Taiwan is thought to rank the outcome of a U.S.–recognized independent state (B) the highest, but to favor capitulation (G) over a forceful invasion by China (H). Finally, when the stakes are low, the U.S. generally would prefer to avoid conflict with China, and favors G over C.

Thus, we make the following minimal assumptions about partial preference orderings:

- $\pi_c(G) \geq \pi_c(E)$: When an unsupported Taiwan does not declare independence, China prefers unification (G) over staying neutral (E).
- $\pi_t(G) \geq \pi_t(H)$: Taiwan prefers to capitulate when China declares war and the U.S. remains neutral.
- $\pi_t(B) > \pi_t(A)$: Taiwan strictly prefers to declare independence if China remains neutral and the U.S. supports Taiwan.

- $\pi_u(G) \geq \pi_u(C)$: The U.S. prefers to avoid conflict with China, given that Taiwan does not declare independence.

2.2 Nash Equilibrium and Stability in the Situation of Taiwan Independence

An outcome is *rational* for a player if she cannot change strategies without decreasing her payoff, holding the other players' strategies fixed. An outcome that is rational for all players is a *Nash equilibrium*; this statement is equivalent to saying that a Nash equilibrium is an outcome from which no player can unilaterally deviate and increase her payoff. Formally, the set of rational outcomes for each player $i = c, t, u$ in Γ is

$$\mathscr{R}_i(\Gamma) \equiv \{s \mid \pi_i(s) \geq \pi_i(s') \ \forall s, s' \in S\}, \tag{1}$$

where s and s' are outcomes in S. The set of Nash equilibria in Γ is

$$EQ(\Gamma) \equiv \mathscr{R}_c(\Gamma) \cap \mathscr{R}_t(\Gamma) \cap \mathscr{R}_u(\Gamma). \tag{2}$$

In the situation of Taiwan independence, outcome A ostensibly is stable, by virtue of being the status quo; we expect A to be contained in the set of Nash equilibria. However, given the weak assumptions made at the end of Section 2.1, if China remains neutral and the U.S. supports Taiwan, then Taiwan prefers to declare independence, upsetting the status quo. Therefore, the status quo cannot be a Nash equilibrium.

On the other hand, outcome G (unification without resistance) is a Nash equilibrium. This is because each player prefers not to change strategies unilaterally when China unifies with Taiwan, Taiwan does not resist, and the U.S. remains neutral.

To resolve this apparent paradox, we re-analyze the situation of Taiwan independence as a three-player metagame.

3 A Metagame Analysis

3.1 Metagame Strategies

Each metagame, $k\Gamma$, is constructed from the base game Γ and is identified by its *title* k, a string of 1, 2, or 3 players, which we denote respectively as k_1, k_2k_1, and $k_3k_2k_1$ (where $k_i \in \{c, t, u\}$ are the players).[2] We refer any players not in the title and their

[2] In actuality, the string may be of any finite length, so that strings with repetitions such as ttcutuuc are allowable. However, as mentioned earlier, Howard (1968) proved that such strings may be collapsed by deleting all but the rightmost appearance of each player (so that ttcutuuc becomes tuc) without affecting the set of metaequilibria.

metagame strategies as being of level-0, while each player k_n and k_n's metagame strategies are of level-i, where $1 \leq i \leq 3$. In the game $c\Gamma$, for example, China is the level-1 player, while Taiwan and the United States are level-0 players. In the game $utc\Gamma$, China is the level-1 player, Taiwan the level-2 player, and the United States the level-3 player. Note that Γ is a special case in which all players are of level-0.

Metagame strategies are:

- for level-0 players, the same as their base game strategies;
- for level-i players ($0 < i \leq 3$), conditional strategies that specify which base game strategy to play for each combination of:

  1) the base game strategies of any other level-0 players,
  2) the metagame strategies of any other level-r players, where $0 < r < i$, and
  3) the base game strategies of any other level-l players, where $i < l$.

For example, consider China's strategies in the metagame $c\Gamma$ (in which China is of level-1, and Taiwan and the United States are of level-0). The strategies of China must specify whether to go to war, *for each combination of the level-0 players' strategies*. An example of one of China's metagame strategies is "*War* if $(NoIndependence, NoSupport)$, but *NoWar* otherwise." Table 2 depicts this metagame strategy.

Table 2 An Example of China's Metagame Strategies in $c\Gamma$.

If		Then
(Taiwan)	(U.S.)	(China)
NoIndependence	*Support*	*NoWar*
NoIndependence	*NoSupport*	*War*
Independence	*Support*	*NoWar*
Independence	*NoSupport*	*NoWar*

Consider another metagame, $uc\Gamma$, in which the U.S. is of level-2, China of level-1, and Taiwan is of level-0. By definition, the level-1 player's metagame strategies specify which base game strategy to play for each combination of the *base game* strategies of the level-0 and level-2 players; thus, China has the same strategies as in 1Γ. The level-2 player's strategies, however, specify which base game strategy to play for each combination of the level-0 player's *base game* strategies, and the level-1 player's *metagame* strategies. The U.S. has $2^{32} \approx 4.3$ billion metagame strategies. An example of one such strategy is, "Support Taiwan only if Taiwan declares independence, and China uses a metagame strategy that specifies not to go to war if Taiwan declares independence."

In $tuc\Gamma$, the metagame strategies of the U.S. and China are the same as in $uc\Gamma$, except that Taiwan's are replaced by ones that specify what do for each combination of the U.S.'s and China's *metagame* strategies. Since there are $2^4 \times 2^5 = 512$ such combinations, the level-3 player (Taiwan) has 2^{512} metagame strategies.

3.2 Solution Concepts in the Metagame

3.2.1 Metagame Strategy Resolution and the Metaequilibrium

The metagame strategy of the highest level player will immediately resolve as a base game strategy. But then, the next highest level player's metagame strategy will also resolve, and also the next next highest, and so on, so that each metagame outcome will always resolves as a base game outcome. Having defined strategies and outcomes in the metagame, we now turn to the concepts of metarational outcomes and metaequilibrium.

A base game outcome s is metarational for player i in $k\Gamma$, if there exists at least one metagame outcome which 1) resolves as s, and 2) is rational for i in $k\Gamma$. A base game outcome s is a metaequilibrium of $k\Gamma$ if it is metarational for all players. Also, it can be proven that if s is a metaequilibrium of $k\Gamma$, then that there exists at least one metagame outcome which 1) resolve as s and 2) is a Nash equilibrium in $k\Gamma$.

3.2.2 Identifying Metaequilibria

The number of outcomes in metagames with three players in the title is astronomical: $2^4 \times 2^{32} \times 2^{512}$. While the Nash equilibria metagames with only player in the title can be found by a brute force search, there are simply too many outcomes to search over when two or more players are in the title. Fortunately, Howard (1971) proved the following theorem, which allows us to identify metaequilibria in the base game without identifying the corresponding Nash equilibria in the metagame.

Theorem 1 (Metarationality Theorem (Howard, 1971)). *For the metagame $k\Gamma$, let L_i equal the set of players to the left of i in the title if i is in the title, or equal the set of all players in the title if i is not in the title. Let R_i equal the set of players to the right of i in the title if i is in the title, or equal the empty set $\emptyset$ if i is not in the title. Let U_i be the set of players not in L_i, R_i, or $\{i\}$. Let S_i be the strategy set of i, and let S_{L_i}, S_{R_i}, and S_{U_i} be the respective joint strategy sets of L_i, R_i, and U_i. Finally, note the base game outcome s^* can be rewritten as $(s^*_{L_i}, s^*_i, s^*_{R_i}, s^*_U) \in S_{L_i} \times S_i \times S_{R_i} \times S_{U_i}$. Then, s^* is metarational in $k\Gamma$ for player i if*

$$\pi_i(s^*) \geq \min_{s_{L_i} \in S_{L_i}} \max_{s_i \in S_i} \min_{s_{R_i} \in S_{R_i}} \pi_i(s_{L_i}, s_i, s_{R_i}, s^*_{U_i}). \tag{3}$$

3.2.3 Symmetric Metaequilibrium

In addition, Howard (1968) proved that if an outcome is metarational in Γ, then it is metarational in $k_1\Gamma$. Furthermore, if an outcome is metarational in $k_1\Gamma$, then it is also metarational in $k_2k_1\Gamma$, and so on, so that metarational outcomes are nested in higher-level games. Since metaequilibria are simply outcomes which are metarational for all players, this implies that the metaequilibria are also nested in higher-level games. For example, the set of metaequilibria of 321Γ contains those in 21Γ, which in turn contains those in 1Γ, which in turn contains those in Γ.

Thus, to find all metaequilibria, we need only look at all games with titles of length three. This idea motivates an additional solution concept, the set of symmetric metaequilibria. A symmetric metaequilibrium is a base game outcome that corresponds to at least one metaequilibrium in $k\Gamma$, for *all* possible titles k. Such an equilibrium is robust in that it is stable for any arrangement of the players' conditional strategy levels.

In the following section, we use (3) to identify symmetric metaequilibria. But since the Metarationality Theorem only identifies metaequilibria, and not their corresponding outcomes in the metagame, we use brute force to find the latter in the metagames with level-1 players. Knowing which higher-level strategy combinations correspond to metaequilibria will, we hope, help elucidate the type of reasoning involved in the metagame.

4 Metaequilibria in the Taiwan independence Game

In this section, we analyze the metagame corresponding to the preference ordering combination (GCHEABFD, BFADECGH, FBAEDHGC). Table 3 lists the metaequilibria for each possible title, while Tables 4–7 show the supporting metagame outcomes for Γ, $c\Gamma$, $t\Gamma$, and $u\Gamma$.

Table 3 Metaequilibria by Title for the Preference Ordering Combination: GCHEABFD, BFADECGH, FBAEDHGC.

Metagame	Metaequilibria
Γ	G
1Γ	BG
2Γ	BG
3Γ	FG
13Γ	BFG
23Γ	BFG
32Γ	BFG
12Γ	ABG
21Γ	ABG
31Γ	BG
132Γ	ABEFG
123Γ	ABEFG
212Γ	ABEFG
231Γ	ABEG
312Γ	ABEG
321Γ	ABEG
Symmetric:	ABEG

Table 4 Equilibria in the Base Game.

Strategies			Equilibrium Outcome
China	Taiwan	U.S.	
War	NoIndep.	NoSupport	$\rightarrow$ G

Table 5 Nash Equilibria in $c\Gamma$ and Corresponding Metaequilibria in the Base Game.

Equilibrium Strategies in the Metagame			Metaequilibrium
China	Taiwan	U.S.	
War if (*NoIndep.*, *Support*) *War* if (*NoIndep.*, *NoSupport*) *Either* if (*Indep.*, *Support*) *War* if (*Indep.*, *Support*)	*NoIndep.*	*NoSupport*	$\rightarrow$ G
Either if (*NoIndep.*, *Support*) *Either* if (*NoIndep.*, *NoSupport*) *NoWar* if (*Indep.*, *Support*) *War* if (*Indep.*, *Support*)	*NoSupport*	*NoIndep.*	$\rightarrow$ G

Table 6 Nash Equilibria in $t\Gamma$ and Corresponding Metaequilibria in the Base Game.

Equilibrium Strategies in the Metagame			Metaequilibrium
China	Taiwan	U.S.	
War	*Either* if (*NoWar*, *Support*) *NoIndep.* if (*NoWar*, *NoSupport*) *Either* if (*War*, *Support*) *NoIndep.* if (*War*, *NoSupport*)	*NoSupport*	$\rightarrow$ G
NoWar	*Indep.* if (*NoWar*, *Support*) *Indep.* if (*NoWar*, *NoSupport*) *NoIndep.* if (*War*, *Support*) *Either* if (*War*, *NoSupport*)	*Support*	$\rightarrow$ B

Table 7 Nash Equilibria in $u\Gamma$ and Corresponding Metaequilibria in the Base Game.

Equilibrium Strategies in the Metagame			Metaequilibrium
China	Taiwan	U.S.	
War	*NoIndep.*	*Either* if (*NoWar*, *NoIndep.*) *Either* if (*NoWar*, *Indep.*) *NoSupport* if (*War*, *NoIndep.*) *NoSupport* if (*War*, *Indep.*)	$\rightarrow$ G
NoWar	*Indep.*	*Either* if (*NoWar*, *NoIndep.*) *NoSupport* if (*NoWar*, *Indep.*) *Either* if (*War*, *NoIndep.*) *Support* if (*War*, *Indep.*)	$\rightarrow$ F

For example, consider $c\Gamma$ (see Table 5), in which China is the level-1 player. The outcome in which Taiwan does not declare independence, the United States does not support Taiwan, and China invades (G) is stable as long as China's metagame strategy *threatens* to invade if Taiwan or the U.S. unilaterally changes their strategy. This is because if Taiwan switches strategies and declares independence, outcome H would result, which Taiwan does not prefer to G. Similarly, if the U.S. switches strategies and supports Taiwan, outcome C would result, which the U.S. does to prefer to G. Of course, since G is China's preferred outcome, China would not change strategies.

On the other hand, the outcome of U.S.-supported Taiwanese independence (B) is stable if China's metagame strategy threatens go to war in the case that the U.S. does not support Taiwan. While this might appear to be a strange equilibrium, it is stable because 1) China cannot improve by going to war, once the Taiwan and the U.S. have jointly used the above strategies, and 2) if China did not threaten to invade, then the U.S. would not support Taiwan, and hence the outcome would not be an equilibrium.

In $u\Gamma$ (see Table 7), G is stable as long as the U.S. threatens not to support Taiwan if Taiwan declares independence. Similarly, Taiwan independence not supported by the U.S. (F) is stable as long as the U.S. threatens to support Taiwan in the case that China goes to war.

Finally, in the metagame $tu\Gamma$, for example, the strategies of China and the U.S. would be of the same form as in Table 7. However, Taiwan's strategies would be of the form "If the U.S. uses the first strategy listed in Table 7 and China chooses *NoWar*, then declare independence; ...," with an if-then statement for every possible combination of the U.S.'s level-1 strategies and China's level-0 strategies.

5 Discussion and Conclusion

The status quo (A) arises as a metaequilibrium in two level-2 metagames, and in all level-3 metagames (see Table 3), lending support to our choice of a metagame analysis of the situation of Taiwan independence.

How robust is the set of metaequilibria we have identified? One approach to answering this question is to consider the general stability of our equilibrium predictions when the assumptions about the preference orderings are changed. To quantify this notion, we define the *signature* of a given set of preference orderings as the 15×8 binary matrix whose columns correspond to the eight outcomes of the game (A-H), and rows to the 15 possible titles (Γ, $t\Gamma$, $u\Gamma$, $c\Gamma$, $tu\Gamma$, $ut\Gamma$, $ct\Gamma$, $tc\Gamma$, $cu\Gamma$, $uc\Gamma$, $tuc\Gamma$, $tcu\Gamma$, $utc\Gamma$, $uct\Gamma$, $cut\Gamma$, $ctu\Gamma$). If an outcome is a metaequilibrium for a given title, the corresponding element in the matrix is defined to be 1; otherwise, the element is defined to be 0.

If the equilibrium prediction we have identified is robust, then changes in the preference orderings should not affect the corresponding signature. Thus, by prespecifying a set of reasonable preference orderings (e.g., based on expert opinion), the relative frequencies of the resultant signatures index their robustness.

In conclusion, we have presented an application of a useful framework for analyzing three-player strategic situations. This framework is suited to the analysis of conflict and cooperation between nation-states, especially when the analyst wishes to make minimal assumptions about the level of reasoning and the preference orderings of players.

References

1. Bicchieri, C. (2006). *The Grammar of Society: the Emergence and Dynamics of Social Norms*. Cambridge University Press, Cambridge, Massachusetts.
2. Fehr, E. and Schmidt, K. (1999). A Theory of Fairness, Competition and Cooperation. *Quarterly Journal of Economics, 114*, 817-868.
3. Howard, N. (1968). *Theory of Metagames*, Ph.D. Thesis , University of London.
4. Howard, N. (1971). *Paradoxes of Rationality*. M.I.T. Press, Cambridge, Massachusetts.
5. Thomas, L. C. (1984). *Games, Theory and Applications*. Dover Publications, Inc., Mineola, New York.