In the Eye of the Beholder?
Cognitive Challenges in Wargame Analysis

Rex Brynen
McGill University
http://www.paxsims.org
what we did

- In September 2017, Jim Wallman and I ran an opportunistic experiment in conjunction with the DIRE STRAITS megagame held at the Connections UK conference.

- Three teams of generally experienced wargamers (with c130+ years of wargame design/facilitation/play experience amongst them):
  - Team USA
  - Team UK
  - Team Euro

- Each team observed the game, was given access to game documentation (briefings, materials produced by the teams, injects and news reporting), and was asked to separately prepare a report on game methodology and findings.
what we found*

- ...the findings were striking. The teams—despite being at the same game—largely disagreed about what they had just seen.

*Caveats to follow.

uh oh...
Pentagon rigged own war games, strategist reveals

Jeff Ballot

FRIDAY, MARCH 11, 2000
UPDATED APRIL 10, 2000

The Pentagon might have anticipated the hit-and-run guerrilla tactics, suicide-bomb attacks and unconventional warfare being waged by the Iraqi regime if it hadn't rigged its own war games last summer.

During that exercise, Paul Van Riper, a newly retired Marine Corps three-star general, repeatedly outlined the Pentagon's best and brightest strategists by impressing, using low-tech communications to issue orders, and thinking outside the box.

When Mr. Van Riper's defending "enemy" Red forces started to thrive at the "attacking" Blue U.S. military, the Pentagon's strategies passed the game, wiped the Red's victory off the board, changed the rules, and reverted to a script that gave the advantage to the Blue attackers.

How We Won the War

By NICHOLAS D. KRISTOF

The American fleet confidently steamed off to war in the Persian Gulf recently -- and promptly got creamed.

This was an elaborate war game, not the real thing, but it reminds us that an invasion of Iraq won't necessarily be a cakewalk. Moreover, a general who participated says that the war game was fiddled with in ways that raise questions about whether the government is returning to a Vietnam-style overoptimism and myopia.

The game, Millennium Challenge 2002, was the largest such simulation ever held, involving 13,500 people. It began, key participants say, with the Americans confidently assuming that they could intercept enemy communications and predict enemy movements.

But the enemy didn't cooperate. It used motorcycle couriers instead of radio and electronic messages, and sent orders as code words inserted into the mezzins' call to prayer -- and this went right by the American intelligence analysts.

The upshot was that the enemy "sank" much of the American fleet as the exercise opened. Oops.
implications?

- Much of our attention has been directed at improving wargame design and facilitation.
- However, there may be serious but largely unexplored challenges in translating a game into findings.
- Specifically, **wargame analysis might depend as much on the analysts as the game.**
  - The findings also suggest that what players learn from an educational game might differ widely from what an instructor thinks players have learned from a game.
- If so, does wargaming have any analytical (or educational) utility at all?
experimental approach (and limits)

- The DIRE STRAITS primarily served an ice-breaker and welcome-to-wargaming function on the first day of CONNECTIONS UK.
- 100+ participants, limited subject matter expertise, no read-ahead
- Megagame format (large, noisy, chaotic, loose kriegsspiel)
experimental bugs and features

- The game methodology not up to the standards of an official/analytical game.
  - HOWEVER, that meant the analytical teams had methodological weaknesses to address.
- US foreign policy behavior in the Trump era—a controversial, even polarizing issue—was an important aspect of game design.
Potential variations in national perspectives were harnessed to explore how preexisting attitudes might affect game interpretation—especially in an era where NATO and FVEY allies increasing view the US as unpredictable/unreliable at best, or a growing political/economic/national security threat at worst.

- POTUS approval rating
  - US: 42% (-11 differential)
  - France 14% (-72)
  - Canada 13% (-57)
  - UK: 11% (-56)
  - Germany 11% (-76)
challenge 1: confirmation bias

- “tendency to search for, interpret, favor, and recall information in a way that confirms one's pre-existing beliefs or hypotheses”
  - Could arise from political views, organizational/doctrinal views, or even views on wargame methodology.
- preferences for early information
  - effects on information searching, assessment, and retention
  - impact of Bayesian updating
  - the role of “motivated beliefs”
challenge 2: framing effects

- How a game looks (name, materials, graphics) can have profound effects on game play—and, presumably, might also have effects on how it is interpreted. (Liberman, Samuels, and Ross 2004)

Both versions were identical games of “prisoner’s dilemma” with identical payoffs.

Reframing the game doubled the rate of cooperative behaviours (with no other changes).

Framing of game had more effect than anticipated player style.
framing effects IN DIRE STRAITS

- DIRE STRAITS used a subgame (rather than a player) to generate US Presidential policy guidance.
- Range of possible US positions, consistent with previous POTUS statements.
- Senior policymakers (SECDEF, SECSTATE, NSA, etc.), political/policy advisors, and loyal military commanders and ambassadors sought to shape US policy outcomes, using a card-based system of policy influence:
famous effects IN DIRE STRAITS

- Policy positions adopted in game closely matched policy statement by POTUS in 2 months following the game.

- But...
framing effects IN DIRE STRAITS

- Tongue-in-cheek graphic presentation of game may have influenced analytical interpretation of US subgame.
  - Team US: subgame broken
  - Team UK: subgame should be more respectful (but OK)
  - Team Euro: subgame seems pretty accurate
groups and analysis

- Group assessment—especially by diverse groups—enhances analytical performance (“the wisdom of crowds”).

  vs

- Groups may tend to lowest-common denominator analysis, or be excessively influenced by senior participants (“groupthink”).

aggregation and team-design methods matter

- DIRE STRAITS experiment suggests some value in having analysts undertake “first impressions” separately, before collective discussion

- equally important to collect player perceptions

- hedgehogs vs foxes (Tetlock 2005)
### Methodology

<table>
<thead>
<tr>
<th></th>
<th>UK</th>
<th>EURO</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>More serious portrayal of Trump Administration needed.</td>
<td>Too many game purposes.</td>
<td>Overdetermined White House game design and adjudication.</td>
<td></td>
</tr>
<tr>
<td>Variable quality form-filling by players.</td>
<td>Untested subgames could generate complications.</td>
<td>Weak/problematic connection between White House game/rest of game.</td>
<td></td>
</tr>
<tr>
<td>More explanation of military status/units/rules needed.</td>
<td>Chinese team too small, and hence overwhelmed.</td>
<td>CONTROL teams inadequately informed of in-game developments.</td>
<td></td>
</tr>
</tbody>
</table>

### Substance

<table>
<thead>
<tr>
<th></th>
<th>UK</th>
<th>EURO</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some nascent conflicts are unlikely to escalate.</td>
<td>Parties like to pull back from the brink in Korea/</td>
<td>Game design precludes meaningful substantive findings.</td>
<td></td>
</tr>
<tr>
<td>Modern 24/7 media causes reactive politics.</td>
<td>Taiwanese willing to engage in risky behavior (but Chinese response was unrealistic).</td>
<td>Game design precludes meaningful substantive findings.</td>
<td></td>
</tr>
<tr>
<td>Cyber operations may be preferred over conventional.</td>
<td>More ASEAN collaboration, less conflict on South China Sea than expected (player behavior was unrealistic).</td>
<td>Game design precludes meaningful substantive findings.</td>
<td></td>
</tr>
<tr>
<td>Diplomatic solutions to North Korea area hard.</td>
<td>More six-party cooperation on Korea than likely in real world (player behavior was unrealistic).</td>
<td>Game design precludes meaningful substantive findings.</td>
<td></td>
</tr>
<tr>
<td>UK, France want to be seen as players—but have little actual power.</td>
<td>Signals from unpredictable/chaotic White House are moderated by foreign policy apparatus, diplomats in the field.</td>
<td>Game design precludes meaningful substantive findings.</td>
<td></td>
</tr>
</tbody>
</table>
Red Teaming game analysis?

- There may be utility in conducting two or more game analyses independently, and then comparing results.

Degree of convergence provides indicator of confidence level.