Wargaming and Forecasting

Rex Brynen
McGill University
http://www.paxsims.org
"wargaming isn’t about prediction"

- Wargames do not provide a solid foundation for high-confidence predictions, because
  - n=1 (usually)
  - Modelling/ V&V challenges
    - Especially outside the kinetic realm
  - Player idiosyncrasies and group dynamics
  - Practical and resource constraints
- Wargamers (and many diplomats, intelligence analysts, and others) are also anxious to deny they do predictions, in case they turn out to be wrong.
- However, wargames are about **plausible futures**: how they might occur, and what effect they might have.
  - Any determination of plausibility is inherently predictive.
but what about “black swans”? 

- Nassim Taleb (2010) on the “ludic fallacy”
  - life isn’t like the neatness of games: there is inadequate information, too much inherent unpredictability, and normal distributions fail to adequately reflect the importance of low probability/high impact events

- ...therefore:
  - prediction is problematic
  - we should build robust, adaptive systems rather than have faith in our ability to anticipate the unanticipated
all swans matter

- DRDC/PCO study of Canadian intelligence forecasts (Mandel, Barnes and Richards 2014) found that analysts' predictions were accurate around 90% of the time.

Figure 5: Reliability Diagram for the MEA Division by Analyst Experience
Central Intelligence Agency, A Tradecraft Primer: Structured Analytic Techniques for Improving Intelligence Analysis (2009)

*Gaming, Modeling, and Simulation are among the more sophisticated techniques taught in a more advanced analytic methods course and usually require substantial commitments of analyst time and corporate resources.
“ersatz history and ersatz people”

- Robert Levine critique (RAND, 1964)
  - games are seductive, and cause players to confuse plausibility with probability
  - players aren’t playing as their real-life counterparts
  - no mechanism for determining relative probability for scenarios
  - Levine argues that crisis games are:
    - cannot test hypotheses
    - poor for determining posture
    - OK as input for planning
    - might work well for other (generally, non-analytical) things

https://www.rand.org/pubs/papers/P7719.html
the challenge of game analysis

- The DIRE STRAITS games at Connections UK 2018 highlighted another problem—that of divergent game analysis.
  - Participants may walk away from the game with different lessons.
  - Analysts may disagree regarding game findings.
### DIRE STRAITS: Methodology

<table>
<thead>
<tr>
<th>METHODOLOGY</th>
<th>UK</th>
<th>EURO</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>More serious portrayal of Trump</td>
<td>Too many purposes.</td>
<td>Overdetermined White House game design and adjudication.</td>
<td></td>
</tr>
<tr>
<td>Administration needed.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variable quality form-filling by players.</td>
<td>Untested subgames could generate</td>
<td>Weak/problematic connection</td>
<td></td>
</tr>
<tr>
<td></td>
<td>complications.</td>
<td>between White House game/rest of</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>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>

### DIRE STRAITS: Substantive Findings

<table>
<thead>
<tr>
<th>SUBSTANCE</th>
<th>UK</th>
<th>EURO</th>
<th>US</th>
</tr>
</thead>
<tbody>
<tr>
<td>Some nascent conflicts are unlikely to</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>escalate.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modern 24/7 media causes reactive</td>
<td>Taiwan willing to engage in risky</td>
<td>Game design precludes meaningful substantive findings.</td>
<td></td>
</tr>
<tr>
<td>politics.</td>
<td>behavior (but Chinese response was</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>unrealistic).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyber operations may be preferred</td>
<td>More ASEAN collaboration, less</td>
<td>Game design precludes meaningful substantive findings.</td>
<td></td>
</tr>
<tr>
<td>over conventional.</td>
<td>conflict on South China Sea than</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>expected (player behavior was</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>unrealistic).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diplomatic solutions to North Korea</td>
<td>More six party cooperation on Korea</td>
<td>Game design precludes meaningful substantive findings.</td>
<td></td>
</tr>
<tr>
<td>are hard.</td>
<td>than likely in real world (player</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>behavior was unrealistic).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UK, France want to be seen as players—</td>
<td>Signals from unpredictable/chaotic</td>
<td>Game design precludes meaningful substantive findings.</td>
<td></td>
</tr>
<tr>
<td>but have little actual power.</td>
<td>White House are moderated by foreign</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>policy apparatus, diplomats in the</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>field.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Twitter Statements on Korea: DIRE STRAITS vs @RealDonaldTrump (August-September 2017)
What does research on forecasting and predictive judgment tell us (Brynen 2017)?
- Cognitive styles and bias matters—and may matter as much, or more, than subject matter expertise (Tetlock 2005; Tetlock and Gardner 2015, Good Judgment Project)
- Crowd-sourcing and aggregation improves forecasting (Tetlock and Gardner 2015)
  - Importance of diversity and team management (Brynen 2017)
- Bayesian updating is important (Silver 2012, Tetlock and Gardner 2015)
- Role-playing improves forecasting, and out-performs “role thinking” (Green 2002, Green and Armstrong, 2011).
- Critical accountability improves performance (Good Judgment Project)
gaming and good forecasting

- Wargames can be a useful adjunct to other forecasting methods, encouraging participants to see the problem in other ways, providing a mechanism for crowd-sourcing, red-teaming assumptions, and encouraging new insights.
- The group dynamics associated with the game may also encourage reframing, network-building, and other desirable ancillary developments.
I found the Iraq simulation an interesting exercise. The methodology might be adaptable for the purposes of thinking about analytical issues related to Iraq. It is a good way to induce analysts to think about aspects of the situation that they might not consider otherwise or look at differently.

However, I wouldn’t see it an alternative to an analytical discussion session, but rather as a preliminary step. The model I would see is a two-part session, starting with 2-3 hours for the simulation and then continuing with another couple of hours of discussion revolving around lessons and insights inspired by the game.
intelligence insights from ISIS CRISIS

The game by itself has **limited analytical usefulness**. Inevitably, players get caught up in the competitive game aspect of the simulation and take whatever steps they can think of to “win,” quibbling about the game mechanics, making unrealistic moves, etc. Some players might get too personally invested in the game and lose sight of the fact that the purpose is to get a deeper understanding of the situation and to think differently than they normally would about it. For example, the reasons put forward for the success or failure of a move are not necessarily well thought-out in the heat of the fray. One possible way around this would be to require the players to list in writing their arguments for and against a move: this would slow things down just a bit and hopefully inspire better arguments. It would also have the advantage of providing a record of the arguments for the later discussion.
The true value of the game would be to provide the basis for a structured discussion afterwards. This could include discussion of several things, including:

- **Insights from players on aspects of the simulation that surprised them**: do they now have a different impression of the strategic situation of one or more sides, of the tactical options available to them (and the restraints they are under), a different understanding of the overall balance of forces, etc.

- **A more considered discussion of the pros and cons for each of the (major) moves during the simulation**: were all of the relevant factors considered, were they given proper weighting, etc. This is where a written record of the arguments would be useful, recorded either by the players at the time and given to the umpire, or compiled by neutral observers.

- **Thoughts from the players on the goals being pursued by their side; their goals may well have shifted as the game progressed**: The introductory notes outlining the situation for each side logically contain general guidance on what that side should be seeking to accomplish, but I must say that in my case I did not pay much attention to the notes and leaped in with my own views on what I wanted to achieve. I think a subsequent discussion of actual and game goals would have been useful.

Overall, a very interesting exercise, and one that has forced me to rethink my previous (rather negative) views of the potential usefulness of simulations in intelligence analysis.
references

- Levine, Robert; Schelling, Thomas; and Jones, William. 1991. Crisis Games 27 Years Later: Plus c'est déjà vu. RAND.