Turning Wargames Into Experiments: Challenges and Approaches

69th Military Operations Research Society Symposium
U.S. Naval Academy
12-14 June 2001

Dr. Daniel R. Tufano, Sr.
POET/Oak Ridge National Laboratory

Dr. Leslie Pierre
Ballistic Missile Defense Organization

Dr. Charles W. Glover
POET/Oak Ridge National Laboratory

This research was partly supported by the US Department of Energy,
Office of Basic Energy Sciences under contract No. DE-AC05-00OR22725 with UT-Battelle, LLC.

Outline

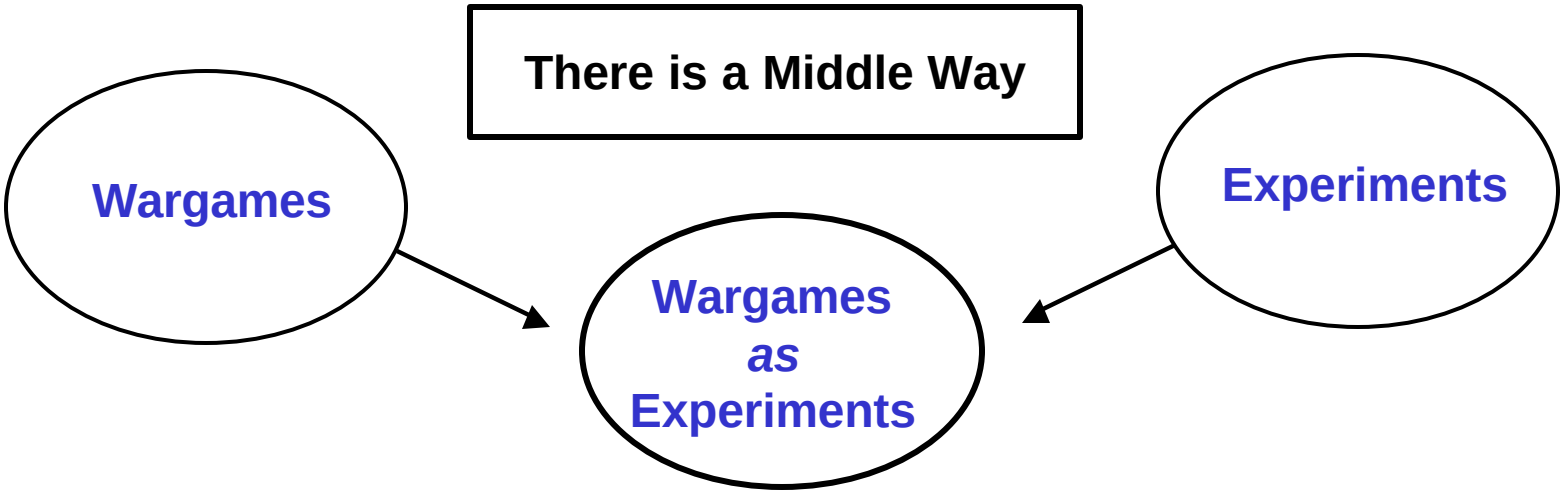
- **Lab Experiments vs. Wargames**
- **Treating Wargames *as* Experiments**
- **Methodology**
- **Data Types**
- **Errors of Interpretation**
- **Summary & Conclusion**

Wargames vs. Experiments

***"War Game Results value has a half life of less than a month"* ***

- **Need more Enduring Results . . . . Knowledge**
  - like lab research

- **But, Lab Controls Can Destroy the Very Phenomena of Interest**



* Perrella, A. M & S Utility Principle. War Game Assessment Workshop; Data Storage, Reduction, and Analysis Session. Joint National Test Facility, 31 July 1998.

Wargames *As* Experiments

"Field Experiments"*

"Field": complex behavioral situations in real or simulated settings

"Experiments": scenarios, participants, conditions, data are focused on a psychological question.

Mental Set of the Investigators: ***Think Experimentally***

- **wargame planning**
- **test design/run matrix**
- **hypothesis generation**
- **data collection/observation plan**
- **analysis & interpretation of results**

* Woods, D. D., Process-Tracing Methods for the Study of Cognition Outside of the Experimental Psychology Laboratory. In [redacted] (Klein, G.A., Orasanu, J., Calderwood, R., and Zsambok, C.E., eds.) Norwood, NJ, 1995. pp. 228-251.

Multi-Player Wargame Constraints

"Had we but the world enough and time . . . . "*

- **All players go through same sequence of test runs**
  - **no counterbalancing of order effects: learning, transfer**
  - **potential confounding variables**

    Fix:
    - **Use same players for repeat games to eliminate learning as factor.**
    - **Plan test run sequence to mitigate transfer effects**

- **One run for each cell in test design matrix: paucity of data**

- **Aggregate measures of effectiveness: hits, leakers, etc.**
  - **one data point per run**

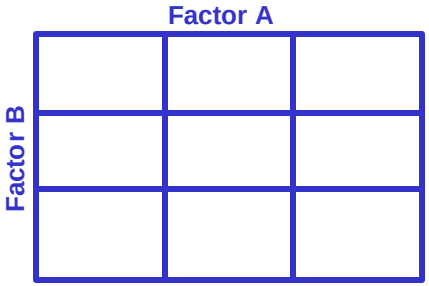
Zero Statistical Power . . . But

* Marvell, A. To His Coy Mistress. ca. 1650-53.

Approach

Plan *as if* you were doing an *experiment.*

- **Test Matrix *looks like* a factorial design.**



- **Generate *a priori* hypotheses *as if* you were going to do (e.g.) an ANOVA:**
  - **main effects**
  - **interactions**

Know what questions you're asking.

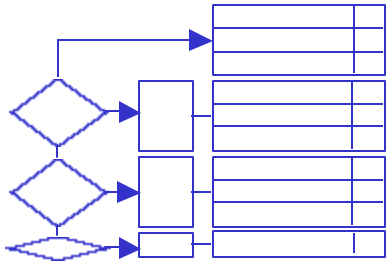
- **Plan convergent data collection methods to fill each cell in the design matrix:**
  - **subjective data from structured but open protocols:**
    - **self-assessment: quantitative & qualitative**
    - **observations**
  - **empirical data reflecting human operator effects**

Let the "story" emerge

Subjective Data

Self Assessment

Rating Scale



Quantitative Data

Diagnostic Questions (1)

"What are you doing?"

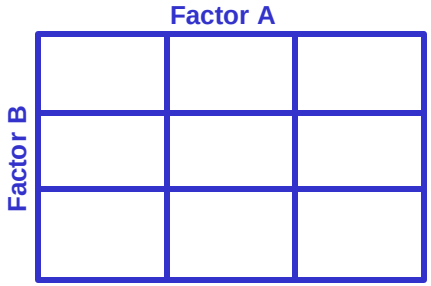
Diagnostic Questions (2)

"What are the consequences?"

Qualitative Data

"Reasons Why"

- **Fill each cell in the design matrix with data:**
  - **numerical**
  - **verbal**

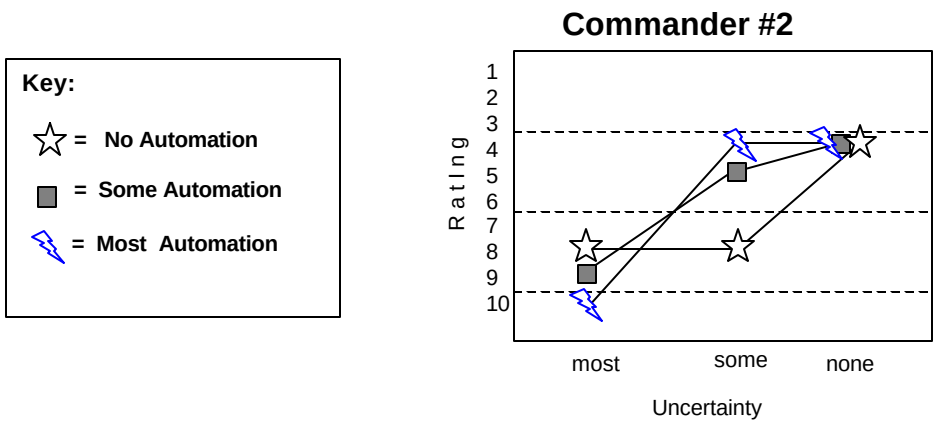


- **Examine Effects**
  - **Main Effects**
  - **Interactions**

- **Interpret Effects with Warfighters' Own Explanations**

Sample Subjective Data

One Individual Warfighter's Data



Examine Data
as if
You Were Doing
an ANOVA

- **Large Uncertainty Main Effect Indicated**
- **Automation X Uncertainty Interaction**

Warfighter Comments:

- **can't deal with degraded tactical picture**
- **need to "control" automation**
- **both together = disaster . . . Can't execute.**

Compare to Other Warfighters' Self Assessments

Empirical Data

Separate Human Effects from Weapons Effects

- **Weapons Effects are Modeled as Fixed p_K values.**
  - **not relevant to human warfighter effects**
  - **random source of variance**

- **Create *Operational Definitions* of the Human "Intangibles"; e.g.:**
  - **situation awareness**
  - **workload**
  - **decision-making ability**

Work from events that are mechanized & digitally logged.

- **Non-Aggregate Measures; 1 per engagement; many per run**

Increase statistical power

- **Measure Actions Leading up to Pulling the Trigger; e.g.:**
  - **decision latency: time between threat detection & weapon target assignment**
  - **decision accuracy: scored by firing doctrine algorithm**

Study the person the bullet.

Analytical Caveats

Know the Sins

Even without inferential statistical tests, errors of interpretation can occur.

- **Type I Error: *falsely rejecting* the null hypothesis; reporting false effects**
  "Factor *x* makes a difference"
  - can be caused by confounds (e.g., uncontrolled order effects)
  - can lead to wasted acquisition $$$

- **Type II Error: *falsely accepting* the null hypothesis**
  "Factor *x* does [redacted] matter"
  - can be caused by lack of statistical power (paucity of data)
  - can be caused by random sources of variance (e.g., weapons effects)
  - can lead to cancelled acquisition of needed capability

We are at greater risk of committing Type II Errors

Yet people think they're being more [redacted] if they only avoid Type I Errors.

Summary & Conclusions

- **Think "Experiment": test plan, design matrix, data collection, analysis**
- **Know the Questions; Generate Hypotheses**
- **Collect Diverse Data on Warfighter Effects, with Convergent Methods**
- **Be aware of [redacted] the Sins, especially Type II Errors**
- **Work Toward Increasing Reliability; But,**
- **Don't Reject the Validity of Warfighters' Data**

Withstand Accusations of *"Heretic"*
by Pharisaical Colleagues