Empirical Research and Decision Making for Engineering Design

Dan Frey
Associate Professor of Mechanical Engineering and Engineering Systems

Outline

- Background
- Research on Design Decision Making
  - Recent critiques of the Pugh Method
  - Our collaborators
  - Model of the first run of the Pugh matrix
  - Model of iterated runs
  - Comparison with alternatives
- Implications for ESD Research?

Popper and Falsificationism

- The criterion of demarcation of empirical science from pseudo science and metaphysics is falsifiability.

- The strength of a theory can be measured by the breadth of experimental results that it precludes

  - Sir Karl Popper , 1934, *Logik der Forschung*

Kuhn and Scientific Revolutions

- Paradigm = a set of scientific and metaphysical beliefs that make up a theoretical framework within which scientific theories can be tested, evaluated, and revised.
- "Normal science" = refinement within a paradigm
- "Revolution" = older paradigm overthrown

The Structure of Scientific Revolutions. Kuhn, 1962

Imre Lakatos and Research Programs

- The scientists involved in a program will attempt to shield the theoretical **core** from falsification attempts behind a protective **belt**

- A **progressive research program** is marked by discovery of novel facts, new experimental techniques, more precise predictions, etc.

- A **degenerating research program** is marked growth of the protective belt in ways that do not lead to novel facts or predictions.

Lakatos, 1978. *The Methodology of Scientific Research Programmes: Philosophical Papers Volume 1*. Cambridge: Cambridge University Pres

The Pugh Matrix

- All comparisons made to a single datum
- Low resolution
- No weighting

Pugh, S., 1990, *Total Design*, Addison-Wesley, Reading, MA.

Discussion Question

- Why was the PuCC chosen as a representation of the early-stage concept evaluation process, instead of other similar methods?

I think that you could make a good case for Kepner-Tregoe instead of PuCC. It was really a matter of personal taste. I think PuCC simplifies the decision process and encourages more creative work.

Pugh Controlled Convergence

- Matrices are run more than once
- Between runs the team works to
  - Generate additional concepts
  - Gather information and analyze alternatives

Pugh, S., 1990, *Total Design*, Addison-Wesley, Reading, MA.

Discussion Question

- Can the PuCC be designed as an automated early-screening process to eliminate large numbers of design concepts before the team of experts enters into the iterative convergence process?

Maybe. My sense is that some filtering may be needed if you have 100 concepts. But it would be easy to go too far. Even if an idea is not likely to "win" overall, it may be good to have in the matrix as a source of hybrids.

What is the Pugh Matrix for?

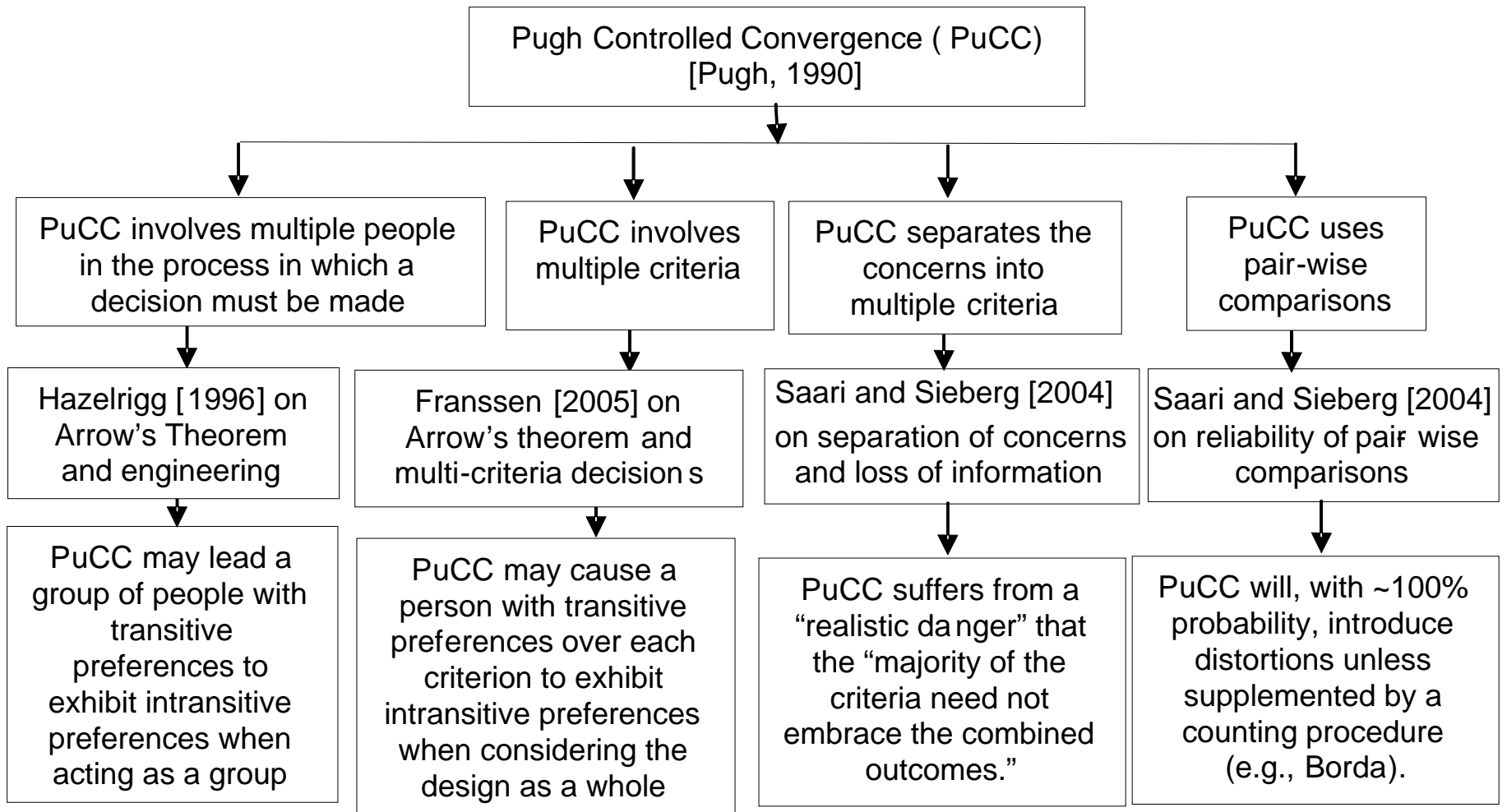The Pugh matrix <u>is</u> for

- Structuring and representing an evaluation procedure
  - Serves as common visual
  - Provides a discipline
  - Helps break down self-sealing behavior
  - Encourages real teamwork
- Convergence
  - Eliminates weaker ideas
  - Retains a set of strong concepts
- Divergence
  - Helps to identify opportunities for combination

The Pugh matrix <u>is NOT</u> for

- Automatic decision making
  - "the scores or numbers … are for guidance only and must not be summed algebraically."
  - "it avoids the rigidity and false confidence of rating/weighting matrices"
- Completely controlling the process
  - "… stimulates creative unconstrained thinking due to its <u>lack of rigorous structure</u>"
- Trade studies
  - More on this today

Pugh, Stuart, 1991, *Total Design*, Addison-Wesley, New York.

Critiques of Pugh's Method

Pugh Controlled Convergence (PuCC)
[Pugh, 1990]

PuCC involves multiple people in the process in which a decision must be made	PuCC involves multiple criteria	PuCC separates the concerns into multiple criteria	PuCC uses pair-wise comparisons
Hazelrigg [1996] on Arrow's Theorem and engineering	Franssen [2005] on Arrow's theorem and multi-criteria decision s	Saari and Sieberg [2004] on separation of concerns and loss of information	Saari and Sieberg [2004] on reliability of pair wise comparisons
PuCC may lead a group of people with transitive preferences to exhibit intransitive preferences when acting as a group	PuCC may cause a person with transitive preferences over each criterion to exhibit intransitive preferences when considering the design as a whole	PuCC suffers from a "realistic da nger" that the "majority of the criteria need not embrace the combined outcomes."	PuCC will, with ~100% probability, introduce distortions unless supplemented by a counting procedure (e.g., Borda).

Franssen, M., 2005, "Arrow's Theorem multi-criteria decision problems and multi-attribute preferences in engineering design," *Research in Engineering Design* 16: 42-56.

Saari, D. G., and K. K. Sieberg, 2004, "Are Partwise Comparisons Reliable?," *Research in Engineering Design* 15: 62-71.

Hazelrigg, G. A., 1996, "The Implications of Arrow's Impossibility Theorem on Approaches to Optimal Engineering Design," *ASME Journal of Mechanical Design*, **118**, pp. 161-164.

Discussion Question

- If the authors knew of Hazelrigg's critique in advance, would their framing of the hypothesis remain the same?

  I think the framing was strongly influenced by Hazelrigg, Saari, and Franssen. I think maybe it's a good thing. If I were a military leader, I would put together a "red team" to pick apart my plans and tactics and I would have to assign members of my own team to the task. In this case, my "red team" is formed at no cost to me, so I'm grateful for it.

Role of the Datum Concept

- The decisions made using Pugh matrices may depend on which datum concept is chosen
- Generally, we are suspicious when a decision can depend on an apparently arbitrary factor (e.g., Arrow's "Independence of Irrelevant Alternatives")
- But are datum concepts really arbitrary?

Discussion Question

- Would it be helpful to evaluate PuCC based on its use in the launch of successful/unsuccessful products, in addition to the computer-based simulation?

Yes, for sure. This is hard to do in practice, but with the right partnerships it should be possible. I expect Chris and I will try hard to do this in the coming years with the Design Center.

Holland's Emphasis on Group Decision Making Persists Today

A sign in the hallways of TU Delft's
faculty of Technology Policy and Management

A View from Cognitive Psychology

Simple heuristics are used by people and animals

Require very little information

Take very little time

Can work remarkably well

Gigerenzer and Todd, G, 1999, *Simple Heuristics that Make us Smart,* Oxford University Press, Oxford, UK.

Research Questions

- Are there merits of PuCC not adequately ackowledged or assessed in the critiques?
- Are there benefits of its simplicity?
  - Low resolution (+, -, S)
  - Comparison to a datum
  - Lack of weighting
- Can such benefits be revealed in a model?
  - Interplay between evaluation and creativity
  - Dynamics of the iterative design process

First Run of the Pugh Matrix

1. Create a set of design concepts to be evaluated
2. Model a set of opinions held by a group of experts
3. Generate the Pugh Matrix
4. Eliminate Concepts Based on the Pugh Matrix

Create a Set of Design Concepts

- The <u>objective merits</u> of each design concept on each criterion are $\mathbf{C}_{ij}$ where $i \in 1 \dots n$ $j \in 1 \dots m$

- Let the datum be assigned index $j=1$
- Sample the criterion scores of the datum
$$\mathbf{C}_{i1} \sim N(s,1)$$

- Sample all other criterion scores
$$\mathbf{C}_{ij} \sim N(0,1), \ j \neq 1$$

defines the strength of the datum concept

Create a Set of Expert Opionions

- The <u>subjective assessment</u> of the criterion score by each expert is

we assume that the experts' opinions are at least partly a function of the objective merits

$$\mathbf{CE}_{ijk} = \mathbf{C}_{ij}\left(1 + \varepsilon_{ijk}\right)$$

the experts' opinions are also affected by other factors which we model as random variables

$$k \in 1 \ldots o$$

$$\varepsilon_{ijk} \sim N(0, \sigma_{ij}^2)$$

defines the degree of variablility of the opinions

Discussion Question

- How does the model's sensitivity to the assumption regarding the independence of expert judgment affect its results? In practice, are there cases where convergence is not likely, given the independence?

I think this independence assumption in our model is a good point for discussion. If the expert's opinions are not just scattered with the "truth" being the average, there is a serious risk. In the limit, the problem is not lack of convergence, but convergence to a very bad design. In my view, the key remedy is that we keep questioning our views critically in the many rounds of evaluation. If any particular judgments are seen to be high leverage, we should test them against models and experiments.

Generate the Pugh Matrix

- The Pugh Matrix entry of each design concept on each criterion is $\mathbf{M}_{ij}$ where $i \in 1 \ldots n$ $j \in 2 \ldots m$

- The entries are

  + if $\mathbf{CE}_{ijk} > \mathbf{CE}_{1jk}$ for all $k \in 1 \ldots o$

 **-** if $\mathbf{CE}_{ijk} < \mathbf{CE}_{1jk}$ for all $k \in 1 \ldots o$

  S otherwise

unaninimity is required as Pugh advocated

Eliminate Concepts

- Pugh proposed no formulaic prescription that automatically leads to the elimination of a concept

- In this model, we eliminate any concept that is <u>dominated</u>

- If another concept in the set is better according to $\mathbf{M}$ along any criterion and is no worse according to $\mathbf{M}$ along any criterion, then the concept will be eliminated

Results from a Model of the First Run of a Pugh Matrix
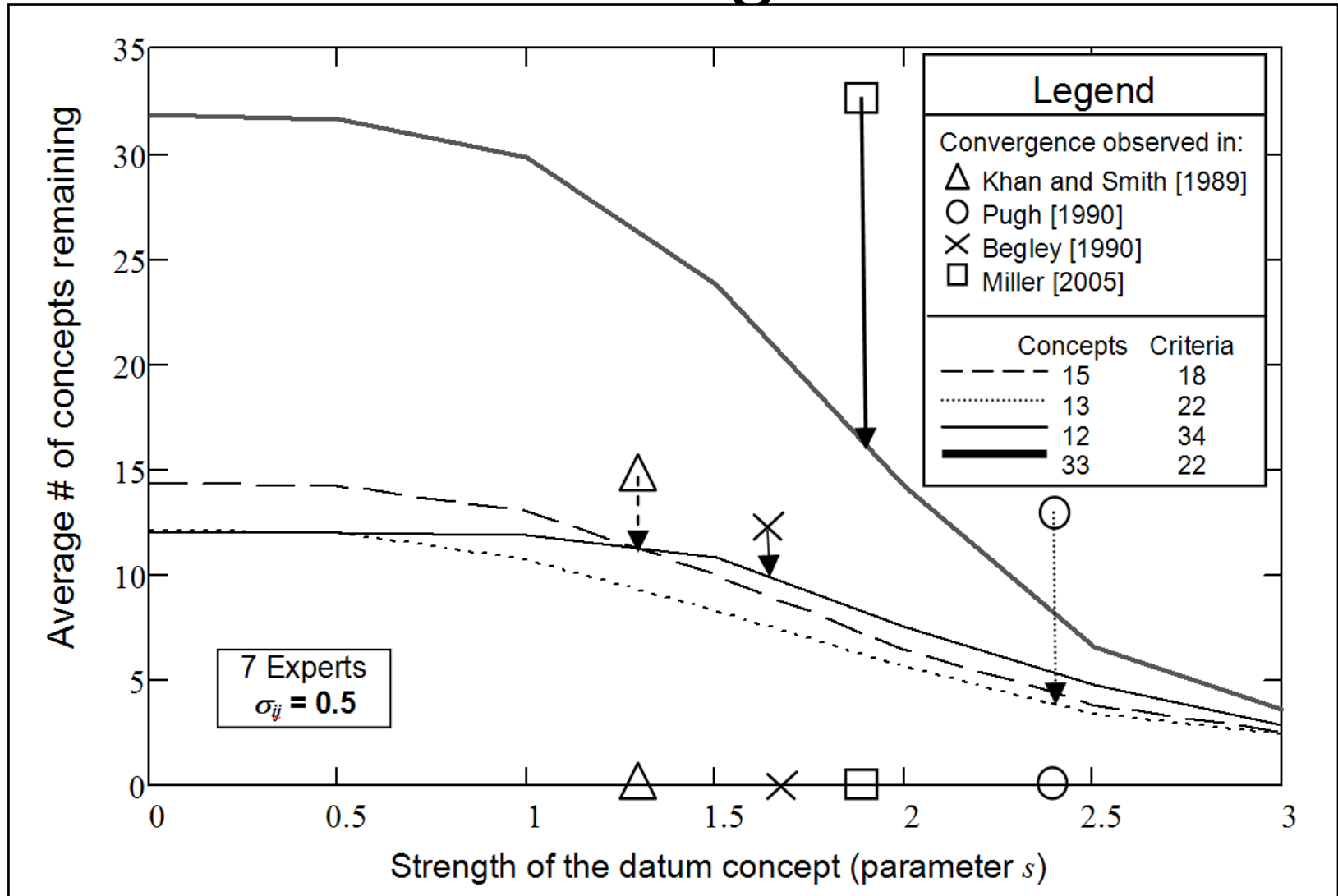
A Model of Design Work Between Matrix Runs

- The $\mathbf{M}$ matrix from the first run is used to guide the process
- Select the top 1/3 of the alternatives based on $\mathbf{M}$
  - Based on the sum of **+**'s and **-**'s
- These are used for
  - Ideation
  - Investigation
  - New datums

A Model of Ideation

- One concept becomes the basis of a hybrid
  - Choose at random from the top 1/3 of the alternatives (according to $\mathbf{M}$)

- Select a second concept that is most complementary to the selected basis
  - Largest number of scores appearing as improvements over the basis (according to $\mathbf{M}$)

- Form a hybrid of the two
  - For each criterion, let the hybrid inherit the larger of the criterion scores $\mathbf{C}_{ij}$ from the its two parents

A Model of Investigation

- For each concept in the top 1/3
- For each criterion wherein it earned an **S** in the previous Pugh matrix
- For each expert, the opinion $\mathbf{CE}_{ijk}$ is refined
  - Reduce the parameter σ_{ij} by a factor of two
  - Replace previous expert opinions by independent samples
- In addition, all the concepts receive a refined estimate in the three most influential criteria
- This is meant to represent the effects of computation, experimentation, interaction with customers, and discussion among the experts
- Investigation moves the criterion estimates of each expert into better alignment with the objective merits

The Convergence of PuCC through Three Iterations With and Without New Concepts Being Generated
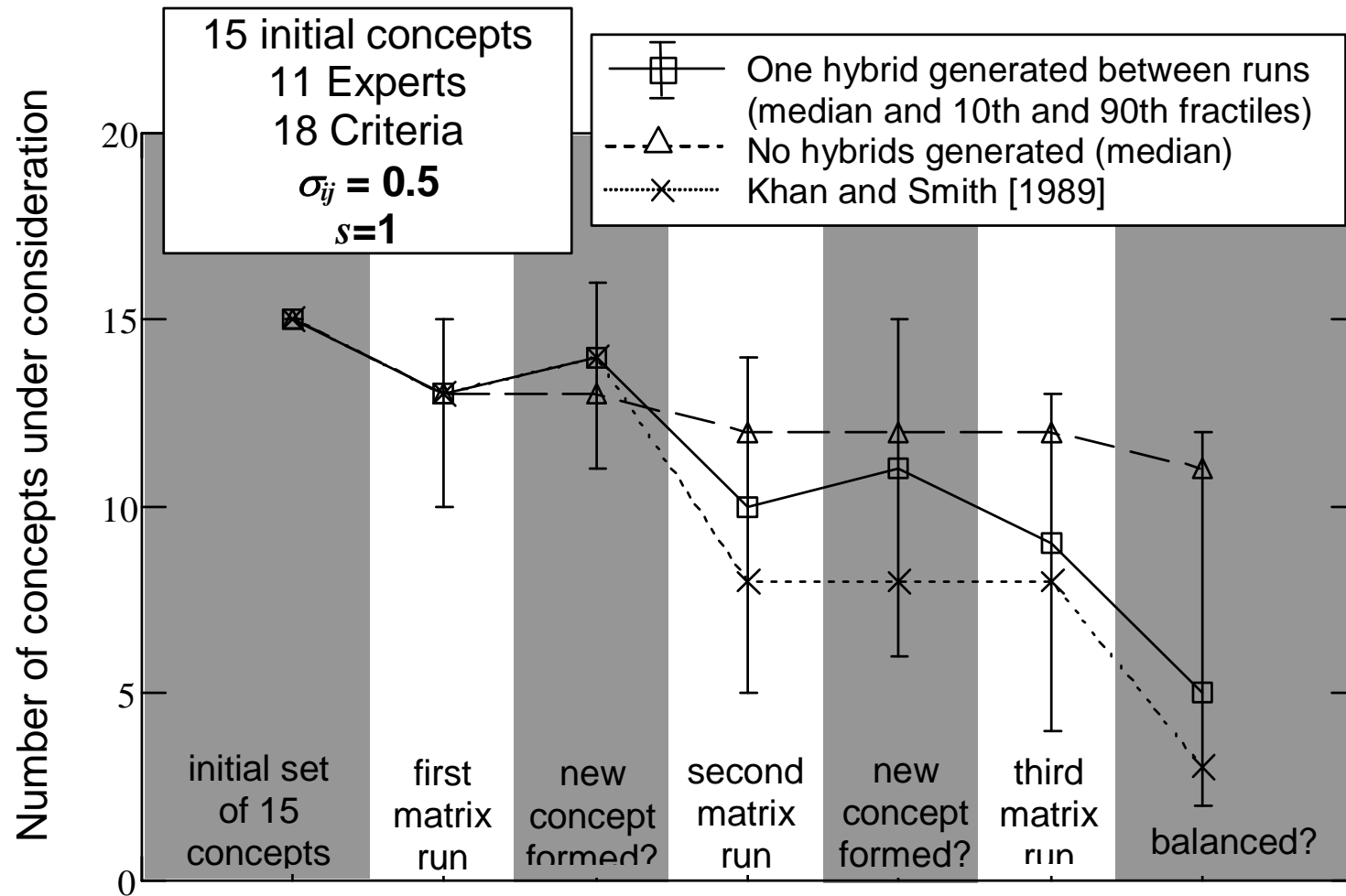
A Model of Profitability

- Let $\mathbf{P}_j$ represent the profit earned by the design, once fully developed and marketed
- Let profit be a function of its objective merits on the evaluation criteria

$$\mathbf{P}_j = \sum_{i=1}^{n} \beta_i \mathbf{C}_{ij} + \sum_{p=1}^{n} \sum_{\substack{q=1 \\ q>p}}^{n} \beta_{pq} \mathbf{C}_{pj} \mathbf{C}_{qj}$$

$$\beta_i \sim \left| N(0,1) \right| \qquad \beta_{pq} \sim N(0, \tau^2)$$

Criteria vary in the degree to which they affect profit, but all are modeled so that larger is better

Parameter τ indicates the degree of difficulties due to separation of concerns

Discussion Question

- How did the authors decide to trade-off simplicity versus complexity, in order to retain the usefulness of the model?

I thought we were favoring simplicity very much. On the other hand, maybe we didn't go far enough. I have so far never been asked to provide a copy of the model so that others can check it, run other scenarios, or for any other purpose. Does that mean it's too complex?

Assessing Designs with a Single Scalar

Image by MIT OpenCourseWare.

Advantage of Pairwise Comparison

- Pugh method uses only "discrimination", never "magnitude estimation"

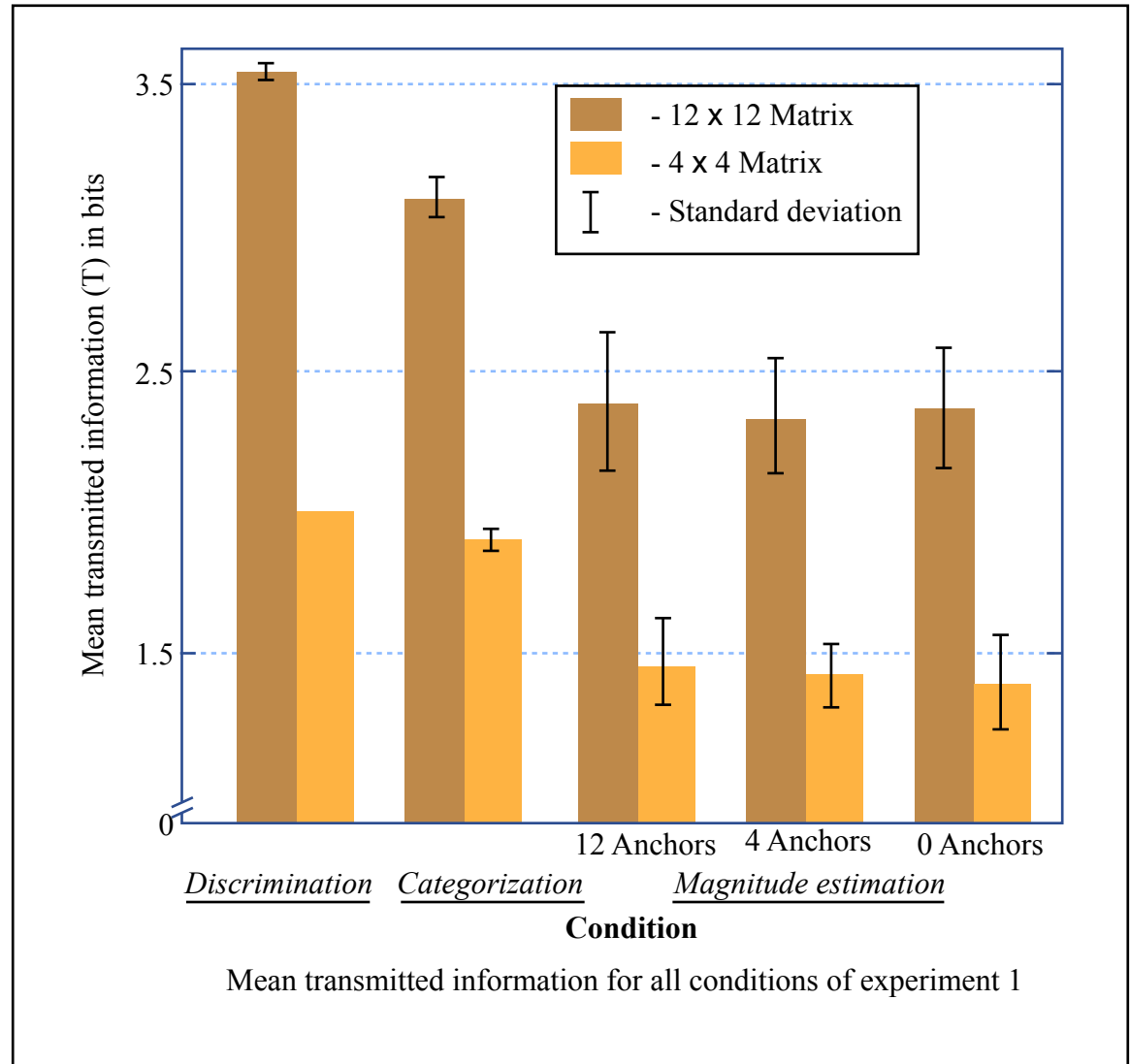- Smith et al. [1984] demonstrated a 2X advantage in accuracy

Mean transmitted information for all conditions of experiment 1

Adapted from: Smith, J, H. Kaufman, and J. Baldasre, 1984, "Direct Estimation Considered within a Comparative Judgment Framework, *American Journal of Psychology* 97(3)343-58.

Conclusions about Pugh's Method

- In response to the critiques of PuCC, we find that the method performs well despite the issues raised by previous authors, especially when
  - Intercriteria interactions are not extremely large
  - A reasonably strong datum can be selected
  - New alternatives are generated

Discussion Question

- If PuCC is not a special case, how would the conclusions change if another early stage design process such as Failure Modes and Effects Analysis (FMEA) or Quality Function Deployment (QFD) were analyzed?

  I think similar assessments of QFD and FMEA can and should be done. In fact, one could make a model of all three being used together since they address different aspects of design. My instinct is that you'd find that a balanced approach using all three is a highly competitive proposition and might be hard to beat.

Conclusions about Design Theory

- Engineering design is not very similar to an election
  - Differences of expert opinion should be resolved more through communication than by majority
  - Apparent conflicts among criteria should not always be handled as trade-offs but instead by "attacking the negatives"
- The interplay of creative work and evaluation is essential
  - Evaluation should guide creative work
  - New alternatives generated in this fashion often greatly simplify decision making

Discussion Questions

- Do the methods of analysis represent a "clash between cultures" in the theory and practice of engineering?

- What are some reasons for only a small percentage of practicing engineers using PuCC?

- Would an ex ante ethnographic study of expert interactions in practice enhance the design and subsequent usefulness of the authors' model, as apposed to ex post experiments and field tests?

- How would an engineering team using the PuCC in practice resolve a divergence or stalemate of experts' opinions, after multiple iterations?

Outline

- Background
- Research on Design Decision Making
    - Recent critiques of the Pugh Method
    - Our collaborators
    - Model of the first run of the Pugh matrix
    - Model of iterated runs
    - Comparison with alternatives
- Implications for ESD Research?

Questions?

ESD.83 Doctoral Seminar in Engineering Systems
Fall 2009