Proceedings of ASME DETC:
Design Engineering Technical Conference
Las Vegas, Nevada, September 4-7, 2007

DETC2007-34758

AN EVALUATION OF THE PUGH CONTROLLED CONVERGENCE METHOD

Daniel D. Frey
Massachusetts Institute of Technology
77 Mass. Ave., Cambridge, MA 02139, USA

Paulien M. Herder
Delft University of Technology
Jaffalaan 5, 2628BX, Delft, the Netherlands

Ype Wijnia
Risk Manager, Essent Netwerk B.V.
Postbus 856, 5201 AW 's-Hertogenbosch
the Netherlands

Eswaran Subrahmanian
Carnegie Mellon University
5000 Forbes Avenue, Hamburg Hall 1209
Pittsburgh, PA 15213, USA

Konstantinos Katsikopoulos
Max Plank Institute for Human Development
Lentzeallee 94, 14195 Berlin

Don P. Clausing
Massachusetts Institute of Technology
77 Mass. Ave., Cambridge, MA 02139, USA

ABSTRACT

This paper evaluates a method known as Pugh Controlled Convergence and its relationship to recent developments in design theory. Computer executable models are proposed simulating a team of people involved in iterated cycles of evaluation, ideation, and investigation. The models suggest that: 1) convergence of the set of design concepts is facilitated by the selection of a strong datum concept; 2) iterated use of an evaluation matrix can facilitate convergence of expert opinion, especially if used to plan investigations conducted between matrix runs; and 3) ideation stimulated by the Pugh matrices can provide large benefits both by improving the set of alternatives and by facilitating convergence. As a basis of comparison, alternatives to Pugh's methods were assessed such as using a single summary criterion or using a Borda count. The models we developed suggest that Pugh's method, under a substantial range of assumptions, results in better design outcomes than those from these alternative procedures.

1. MOTIVATION

We suggest there is a dissonance between statements in the design theory and methodology literature and observations of engineering practice. Recent papers have described current engineering practice as seriously deficient in many central aspects of design:

- "Multi-criteria decision problems are still left largely unaddressed in engineering design." [Franssen, 2005]
- "A standard way to make decisions is to use pairwise comparisons …Pairwise comparisons can generate misleading conclusions by introducing significant errors into the decision process … rather than rare, these problems arise with an alarmingly high likelihood." [Saari, 2004]
- "...Engineering design can be performed without resort to the axiomatic framework ... But there will be an attendant loss, and in my experience this loss is typically a factor of two or more in the bottom line, such as profitability." [Hazelrigg, 1999]

On the other hand, over the past century, engineering has transformed transportation, housing, communication, sanitation, food supply, health care, and almost every other aspect of human life [Constable and Somerville, 2003]. Studies of economic growth suggest that technical innovation accounts for more than 80% of long term improvement [Solow, 1957]. There are opportunities to improve engineering practice, but such improvements require an understanding of how the methods used today contributed engineering's remarkable record of success. More specifically, in contrast to the quotes listed above, we submit that:

- Multi-criteria decision problems are, with methods in use today, addressed successfully in engineering design. The benefits of considering multiple criteria (rather than summary measures only) outweigh the potential risks.
- Pairwise comparison has substantial benefits because it is psychologically easier than other procedures. To the extent that serious errors are made in engineering, they are usually due to other factors rather than information losses due to pairwise comparison.
- Most engineering firms would not improve profitability by adopting an axiomatic framework for decision making. The resources required to implement an axiomatic approach would negatively affect more important activities and the net effect would be reduced profitability.

This paper seeks to contribute to exploring these hypotheses by analyzing a design method, Pugh Controlled Convergence, and by analyzing its relationship to recent developments in design theory. Pugh's method is implicated either explicitly or implicitly in several recent developments. Figure 1 illustrates how Pugh Controlled Convergence has been subject to critique. In the second layer of the diagram, we list some features of Pugh's method. Below that, we list papers whose critiques of Pugh's method result from those features. In the bottom row of the Figure 1, we list aspects of our model that are meant to help respond to each critique.

Copyright © 2007 by ASME

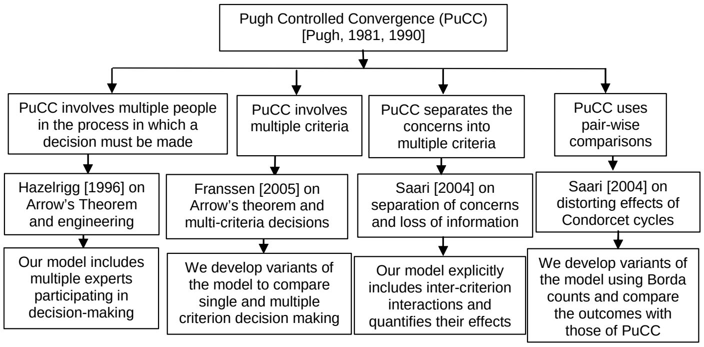
Pugh Controlled Convergence (PuCC) [Pugh, 1981, 1990]
PuCC involves multiple people in the process in which a decision must be made
PuCC involves multiple criteria
PuCC separates the concerns into multiple criteria
PuCC uses pair-wise comparisons
Hazelrigg [1996] on Arrow's Theorem and engineering
Franssen [2005] on Arrow's theorem and multi-criteria decisions
Saari [2004] on separation of concerns and loss of information
Saari [2004] on distorting effects of Condorcet cycles
Our model includes multiple experts participating in decision-making
We develop variants of the model to compare single and multiple criterion decision making
Our model explicitly includes inter-criterion interactions and quantifies their effects
We develop variants of the model using Borda counts and compare the outcomes with those of PuCC

Figure 1: Features of Pugh's method, critiques related to each feature, and our model-based approach to testing those claims.

In section 2, we discuss the recent papers on this topic and the implications for Pugh's method. In sections 3 and 4, we build and explore a model of the design process. Using the framework described by Frey and Dym [2006] we construct computer executable entities meant to represent, in abstract form, the aspects we consider most essential to understand Pugh Controlled Convergence. Our model explicitly includes: 1) the role of the datum concept, 2) the convergence of expert opinion based on investigation, and 3) the generation of new alternatives. These considerations have not played a prominent role in the scholarly debate on design decision making, but it seems to us that they have a first order impact in practice. In light of these considerations, we seek to ascertain whether or not the reported undesirable behaviors of Pugh's method actually arise under realistic conditions. If they do arise, we seek to determine their severity and suggest what conditions are required to avoid them.

2. BACKGROUND

2.1 Review of Pugh Controlled Convergence

As background to our model-based evaluation, we review the method Stuart Pugh developed. Pugh stated that after market investigation, product design specification, and concept synthesis, the design team should engage in an iterative process of culling down and adding to the set of concepts. We will refer to the method as "Pugh Controlled Convergence" or PuCC. The goals of the process are: 1) to converge on a strong concept that has promise of out-competing the current market leader; and 2) to help the team understand the reasons for their choice.

A prominent aspect of that process is presentation and discussion of information in the form of a matrix. The columns of the matrix are labeled with a description, in drawings and text, of design concepts. The rows of the matrix are labeled with concise statements of the criteria by which the design concepts can be judged. The method requires selection of a datum, preferably a design concept that is both well understood and known to be generally strong. Often the datum concept is currently the leader in the market. Evaluations are developed and entered into the matrix through a facilitated discussion among the experts. Each cell in the matrix contains symbols **+**, **-**, or **S** indicating that the design concept related to that column is clearly better than, clearly worse than, or roughly the same as the datum concept as judged according to the criterion of that row.

Figure 2 depicts the matrix created by a design team that used Pugh's method. Academic publications on the method will often convert such results into neatly formatted tables, but this may contribute to a misunderstanding of what is actually done. As Figure 2 suggests, the PuCC process is simple and coarse-grained. Observation of teams carrying out the process show the method is also flexible and heuristic. We assert that these are affirmative benefits, making the method fit well into the early stages of concept design. For example, alternatives to Pugh's method often require greater resolution of the scale (suggesting five or ten levels rather than just three) and often require numerical weighting factors. Pugh found that this sort of precision is not well suited to concept design. Practical experience may support this position, but we will not try to make that case here. Instead, a model-based analysis will be used to evaluate the hypothesis that a simpler method with a coarse scale gives better results.

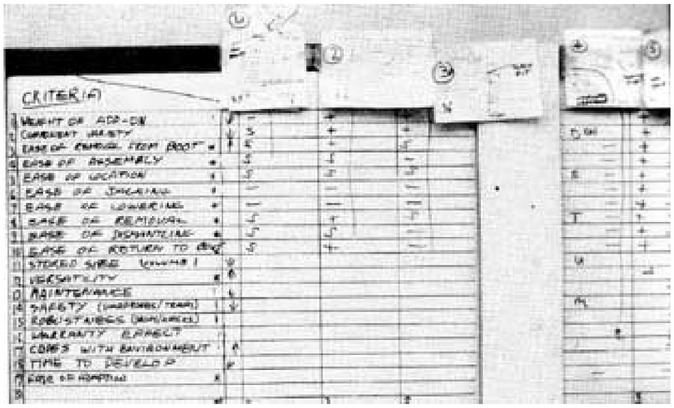

Figure 2. A physical manifestation of the Pugh matrix.

It is important to note that there is no voting in Pugh's method. Let us consider a situation in which several experts claim that a concept is better than the datum and others disagree. In Pugh's method, a discussion proceeds in which the experts on both sides communicate their reasons for holding their views. In many cases, this resolves the issue because either: 1) facts are brought to light that some individual experts did not previously know, 2) a clarification is made about what a design concept actually entails, or 3) a clarification is made about what the criterion actually means. If that discussion leads to an agreement among the experts, then a **+** or **-** may be entered in the matrix. If the disagreement persists for any significant length of time, then an **S** is entered in the cell of the evaluation matrix. In Pugh's method, **S** can denote two different situations. It can mean that the experts agree that the concept's merit is similar to the datum or that the differences between the concept and the datum are controversial and cannot be determined yet. In this case, team members would be encouraged to find additional information necessary to resolve the difference of opinion (Pahl and Beitz [1984] have suggested an "**i**" or "**?**" should be entered to more strongly encourage investigation).

Generally, the evaluation matrix includes summary scores along the bottom. The number of **+**, **-**, or **S** scores for each concept are counted and presented as a rough measure of the characteristics of each alternative. This raises an important issue. These scores are sometimes interpreted as a means by which to choose the single winning design. This misconception is reflected in terminology -- Pugh's method is most often referred to in the design literature as "Pugh Concept Selection" whereas Pugh emphasized "Controlled Convergence." The term "Concept Selection" would seem to imply that after running a matrix a single alternative will be chosen. This is not an accurate characterization of the PuCC process. The first run of the evaluation matrix can reduce the number of design concepts under consideration, but rarely reduces the set to a single alternative. A matrix run can result in at least four kinds of decisions (not mutually exclusive) including decisions to: 1) eliminate certain weak concepts from consideration, 2) invest in further development of some concepts, 3) invest in information gathering, and 4) develop additional concepts based on what has been revealed through the matrix and the discussions it catalyzed. To follow up on these actions, the matrix should be run iteratively as part of a convergence process.

To illustrate how iterated runs of the evaluation matrix result in convergence, consider a real-world example. Khan and Smith [1989] describe a case in which a team in industry designed a dynamically tuned gyroscope. The process began with 15 design concepts and 18 criteria, which we would characterize as a typical problem scale.

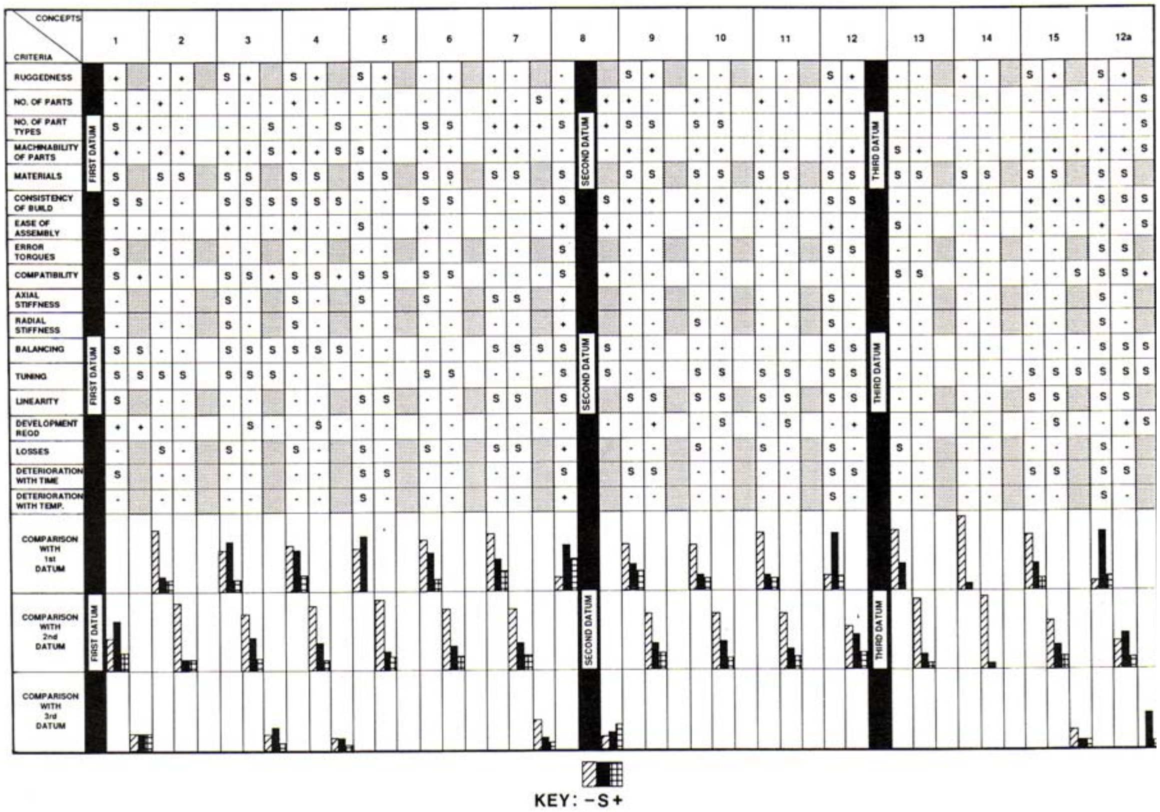

Figure 3. Data from three runs of Pugh matrices in the design of a gyroscope (from Khan and Smith [1989]).

Figure 3 depicts three runs of a Pugh matrix each with a different datum concept. In this case, the process provided little convergence in the first round with no concepts eliminated (although perhaps concepts numbered 5 and 13 might have been removed since they were dominated). After the second run of the matrix the set converged to eight concepts. The third run enabled the team to decide on a single alternative, but notably the concept chosen (labeled 12a) was not even present in the initial set of concepts considered but rather emerged as a hybrid of two others.

As the case study by Khan and Smith [1989] shows, the PuCC process includes decision making, but it cannot be sufficiently modeled *only* as decision making. The process involves learning and creative synthesis also and there is no bright line when these activities stop and decision making begins. Learning, synthesis, and decision-making proceed in parallel and synergistically. The analysis and discussion of design concepts catalyzes creation of additional concepts, which in turn may simplify decision-making. This interplay among decision-making and creative work is often neglected when considering the merits of design methodologies. Our models in Sections 3 and 4 explicitly include these aspects of the design process.

The Pugh method is quite well known, but it is not clear how widely it is used. One survey of Finnish industry suggested Pugh's method is used by roughly 2% of firms [Salonen and Pertutula, 2005]. Informal approaches labeled as "concept review meetings", "intuitive selection" or "expert assessment" are dominant, all estimated to be used in about 40% of cases. These data, although not conclusive, suggest that formal design methods are generally under-utilized. We wish to present a case for methods that provide an appropriate degree of structure, neither too much nor too little. Later sections of this paper are addressed to this objective. First we review some literature that presents technical objections to Pugh Controlled Convergence.

2.2 Pugh, Utility, and Arrow's Theorem

Hazelrigg [1998] has proposed a framework for Decision-Based Design (DBD). A central feature of the framework is that the choice among alternative designs is impacted by the decision maker's values, subjective uncertainties, and economic factors such as demand at a chosen price. Hazelrigg's DBD framework requires rolling up all these diverse considerations into a single scalar value -- utility as defined by von Neumann and Morgenstern [1953]. Having computed this value for each alternative configuration, the choice among the design alternatives is simple -- "the preferred choice is the alternative (or lottery) that has the highest expected utility" [Hazelrigg, 1999].

Although Hazelrigg's framework for DBD is subject to much debate, it continues to have significant influence in the community of researchers in engineering design. The textbook *Decision Making in Engineering Design* [Lewis, Chen, and Schmidt eds., 2006] reflects a wide array of opinions on how decision theory can be implemented in engineering design, but also demonstrates that the core ideas of the DBD framework are being developed actively.

Hazelrigg's framework explicitly excludes the use of Pugh's method of Controlled Convergence. Hazelrigg states the conclusion in broad terms explaining that the acceptance of von Neumann and Morgenstern's axioms leads to one and only one valid measure of worth for design options. Since Pugh's method does not explicitly involve computation of utility, Hazelrigg has argued that Pugh's method is invalid. Also, DBD invokes Arrow's General Possibility Theorem [Arrow, 1951]. Hazelrigg [1999] states "in a case with more than two decision makers or in a multi-attribute selection with more than two attributes, seeking a choice between more than two alternatives, essentially all decision-making methods are flawed."

Scott and Antonsson [2000] argue that the implications of Arrow's theorem in engineering design are not nearly so severe. A principal basis for this conclusion is that "the foundation of many

Copyright © 2007 by ASME

engineering decision methods is the explicit comparison of degrees of preference." This line of approach to the possibility of choice is similar to Sen's who states "Do Arrow's impossibility, and related results, go away with the use of interpersonal comparisons ...? The answer briefly is yes" [Sen, 1998]. In combining the influence of multiple attributes, Scott and Antonsson state that "there is always a well-defined aggregated order among alternatives, which is available to anyone with the time and resources to query a decision maker about all possible combinations." The DBD framework establishes the aggregated order via expected utility, but Scott and Antonsson concluded that "the relative complexity of these methods is not justified" compared to simpler procedures such as forming using a weighted arithmetic mean. Pugh's method represents a further simplification and this paper seeks to determine whether this additional reduction in complexity is also justified.

Franssen [2005] attempted to counter the arguments by Scott and Antonnsen. Franssen challenges, on measure theoretic grounds, the existence of a global preference order that is determined by any aggregation of individual criterion preference values. Franssen argues that if criterion values are ordinal or interval, then the global aggregated order posited by Scott and Antonsson cannot be defined or else that it will be subject to Arrow's result. More fundamental however, is Franssen's assumption that measurable attributes of the design can never determine the designer's overall preference ordering. Franssen holds that "it is of paramount importance to realize that preference is a mental concept and is neither logically nor causally determined by the physical characteristics of a design option." Franssen concluded that "Arrow's theorem applies fully to multi-criteria decision problems as they occur in engineering design." Franssen also draws specific conclusions regarding Pugh's method:

> *... This method ... can attach different global preferences, depending on what is taken as the datum ... Hence it does not meet Arrow's requirement ... It is important not to be mistaken about what Arrow's theorem tells us with respect to the problem. ... What it says is that, for any procedure of a functional form that is used to arrive at a collective or global order, there are specific cases in which it will fail .... Accordingly, for any specific procedure applied, one must always be sensitive to the possibility of such failures.*

This quote by Frannsen is a major motivation for this paper. Our model-based assessment of Pugh's method of controlled convergence will explicitly deal with the issue that the selection of the datum does make a difference in running the matrix (although the matrix is typically run multiple times with different datum concepts). And, as Franssen notes, one must always be sensitive to the possibility of failures induced by one's chosen design methods. But the *possibility* of failure is not enough to justify abandoning a technique that has been useful in the past. This paper seeks to quantify the impacts of such failures and weigh them against the benefits of the PuCC process.

2.3 Pugh and Pairwise Comparison

Saari [2004] constructed an argument against all uses of pairwise comparisons in engineering design except for very restricted classes of procedures. Going beyond the argument based on Arrow's theorem which only claims the *possibility* of error, Saari makes specific claims about the *likelihood* and *severity* of the errors. Saari proposes a theorem including the statement that "it is with probability zero that a data set is free from the distorting influence of the Condorcet n-tuple data." From this mathematical statement he draws the practical conclusion that pairwise comparisons "can generate misleading conclusions by introducing significant errors into the decision process … rather than rare, these problems arise with an alarmingly high likelihood."

Saari's claims that "even unanimity data is adversely influenced by components in the Condorcet cyclic direction." In Pugh's method, designs that are unanimously judged to be superior across all criteria will never be eliminated. Therefore the distorting effect is not always reflected in the alternative chosen, but in some other regard. Saari states "suppose the $A \succ B \succ C$ ranking holds over all criteria ...If we just rely on the pairwise outcomes, this tally suggests that the $A \succ B$ and $A \succ C$ rankings have the same intensity... It is this useful intensity information that pairwise comparisons lose...". This raises an important point related to intensity of feelings. It is not enough that an engineering method should lead to selection of a good concept. It is also essential that the method should give the team members an appropriate degree of confidence in their choice. But Saari's proposed mathematical processing of the team members' subjective opinions may not have the desired result. We suggest that a psychological commitment to the decision may be attained more effectively by convergence of opinion rather than balancing opinions as if design were an election. As differences of opinion are revealed by the Pugh process, investigation and discussion ensue. Since we consider this an important part of engineering design, we seek to incorporate in our model the possibility that people can discover objective facts and change their minds.

A second theme in Saari's paper regards separation of concerns. Pugh's method explicitly asks decision makers to consider multiple criteria by which the options might be judged. Saari claims that this separation of the information leads to a "realistic danger" that the "majority of the criteria need not embrace the combined outcomes." Saari's argument for this conclusion is "Engineering decisions often are linked in the sense that the $\{A,B\}$ outcome is to be combined with the $\{C,D\}$ conclusion. For instance, a customer survey may have $\{A,B\}$ as the two alternatives for a car's body style while $\{C,D\}$ are alternative choices for engine performance." Saari then outlines an imaginary scenario in which the survey data lead to a preference reversal due to an interaction among criteria. The survey data in the scenario suggest that although customers prefer body style A when considered separately and engine performance C when considered separately, they do not prefer the combination of those particular body styles and engine performance options. Saari concludes the resulting product "runs the risk of commercial failure" and that "product design decisions ... could be inferior or even disastrous."

With the argument regarding separation of concerns, Saari may have sacrificed his claim that these events occur with high likelihood. Many inter-criterion interactions in engineering are known *a priori* to be too small to cause the reversals Saari describes. Consider a specific example in which a team designed a gyroscope and needed to consider criteria such as "machinability of parts" and "axial stiffness" [Khan and Smith, 1989]. The sort event that Saari asks us to consider is that design concept is better than another on "machinability of parts" and better on "axial stiffness" but that the ways those criteria combine in the design concepts makes them inferior jointly. It seems unlikely to us that preference for more axial stiffness can be reversed by the machinability of the parts by which it is composed. On the other hand, there are examples one can construct in which such inter-criterion interactions seem more likely, especially regarding aesthetic facets of a design. But in the context of Pugh's method, we need to consider whether these interactions actually lead to choice of weak concepts. This is a motivation for the models we will introduce in section 4.1.

The analysis by Saari is not only a warning regarding potential risks, but is also presented as a guide to modifying the design process -- "Once it is understood what kind of information is lost, alternative decision approaches can be designed." Unfortunately, the proposed remedies impose significant demands on information gathering and/or processing. Saari suggests a procedure involving "adding the scores each alternative gets over all pairwise comparisons." Let us consider what this implies for the Pugh process using the specific example in Khan and Smith [1989]. The process began with 15 design concepts and 18 criteria. The first run of the matrix therefore demanded that 14 concepts be compared with the datum across 18 criteria so that 252 pairwise comparisons had to be made by the team to fill out the first evaluation matrix. If the run of the matrix was to be completed in a standard 8-hour work day, then about 2 minutes on average could be spent by the team deliberating on what symbol should be assigned to each cell. In reality, many of the cells might be decided upon very quickly because the difference between the concept and the datum is obvious to all concerned. However, even accounting for this, the time

Copyright © 2007 by ASME

pressures are quite severe. Saari's remedy requires that every possible pairwise comparison must be made requiring 15 choose 2 pairwise combinations of concepts across 18 criteria -- 1890 pairwise comparisons in all. If the process is to be completed in a single work day, there would be only 15 seconds on average per comparison. Alternately, one might preserve the same average discussion time per cell (2 minutes) and allow around 63 working hours for the task rather than 8. Given this expansion of resource requirements, it is possible Saari's suggested remedy is more harmful than the Condorcet cycles themselves. We seek to consider this hypothesis quantitatively in section 4.3.

2.4 Pugh and Rating, Weighting, and Sensitivity

Takai and Ishii [2004] presented an analysis of Pugh's method including comparison with alternative approaches. The paper articulates three desiderata of concept evaluation methods: 1) The capability to select the most preferred concept, 2) The capability to indicate how well the most preferred concept will eventually satisfy the target requirements, and 3) The capability to perform sensitivity analysis of the most preferred concept to further concept improvement efforts.

To evaluate the Pugh method, Takai and Ishii suggest three possible modifications of Pugh's matrix. Two of the modifications involve types of rating and weighting. One of the modifications involves computing the probability of satisfying targets. In a case study involving design of an injector for a new linear collider, they consider the merits of three alternatives over nine criteria. All four methods suggested the same design as the most preferred alternative. However, a further analysis suggested that even the most preferred concept had only an 8.9% chance of satisfying its requirements and that if availability were improved by 3% and cost reduced by 30%, then the chances of success improved to 76.8%. They conclude that the approach involving quantifying one's beliefs in terms of distributions, evaluating concepts by probability of satisfying targets, and performing sensitivity analysis is the most promising approach.

The analysis by Takai and Ishii seems appropriate to situations in which the number of alternatives is small, all the alternatives are well characterized, and the possibility of generating new concepts is not available. Such a scenario is likely to arise at some stage in the convergence process, but perhaps such modifications are counterproductive in the earlier stages. If probabilistic analyses were conducted with rather coarse estimates, there may be a risk of misleading the team into false confidence. Pugh and Smith [1976] argue that numbers used in evaluation matrices are easily interpreted as similar in standing to the sorts of objective number engineers most often work with (e.g., densities, voltages, and elastic moduli). Overly precise representations create a risk of unwarranted faith in decisions based on rough estimates. It is possible that, in the early stages of design, the same time and resources needed for probabilistic analysis might be used in some more productive way. The model we propose in Section 3 is intended to enable exploration of such trade offs among different emphases and different styles of work.

2.5 Review of Relevant Psychology Research

To address the various critiques and the proposed improvements of PuCC, it is worthwhile to review some results from psychological research. The discipline of psychology can provide insight into what is and is not possible for humans to do or to understand. Psychology also provides information about human capacities that can be leveraged by decision making methods. This section reviews selected topics helpful to understanding later parts of this paper.

Decision Field Theory (DFT) is an approach to modeling human decision making. The theory acknowledges that humans make decisions by a process of deliberation which is inherently dynamic with degrees of preference varying over time [Johnson and Busemeyer, 2005]. DFT models can be created that simultaneously accord with a large set of empirically demonstrated effects and has been used to analyze a variety of decision tasks including, most relevant to engineering, multi-attribute decision making under time constraints [Diederich, 1997]. The models described in Section 3 bear some resemblance to those from Decision Field Theory since they are dynamic with states varying through repeated cycles based on previous states. A difference of our approach from DFT is that we do not model decision making as emerging from weighting of valences primarily, but instead model decision making as determined by decision rules or heuristics. Psychology research has shown that such heuristics are often more robust than schemes involving weighting, especially in generalizing from experience to new tasks [Czerlinski, Gigerenzer, and Goldstein, 1999].

This paper also relies upon some research on human perception and judgment. Psychologists draw a distinction among discrimination and magnitude estimation. In a discrimination task, a human subject is asked to compare two entities and decide which has a property to a greater degree. In a magnitude estimation task, a human subject is asked to give a quantitative value for an entity along a continuous scale. Smith et al. [1984] conducted a study in which human subjects were asked to make judgments about line length under various task conditions. The study showed that human judgment is much less prone to failure (by roughly a factor of two) when two entities are compared directly rather than estimating values on a continuous scale. We suggest that the study by Smith et al. [1984] demonstrates an affirmative value of paired comparison which may be relevant to Pugh's method. By exclusively using comparisons with a datum and avoiding rating and weighting, the judgments made by engineers may be less prone to error. The implications of this hypothesis will be explored in Section 4.

3. A MODEL OF PUGH CONTROLLED CONVERGENCE

This section presents a quantitative model of the Pugh Controlled Convergence process. The model is a highly abstract representation of the process we have observed among real teams using the method. It is important to keep in mind that "essentially, all models are wrong, but some are useful" [Box and Draper, 1987]. Although this model cannot hope to capture, in all its facets, how concept design actually proceeds, we envision that people can use the model to probe their beliefs about decision-making and its role in engineering design.

3.1 A Model of the First Round of the Evaluation Matrix

This section describes a basic model of the first round of an evaluation matrix. The model is stochastic, so the model is executed in many independent trials so that we can characterize the behavior statistically. In each trial, the simulation is comprised of the following four steps:

1. **Create a set of design concepts to be evaluated.** In the model, there are values $\mathbf{C}_{ij}$ where $i \in 1 \ldots n$ and $j \in 1 \ldots m$. Each value $\mathbf{C}_{ij}$ represents the objective merit of concept *j* on criterion *i*. These objective merits will influence the Pugh matrix, but the two matrices are not equivalent since $\mathbf{C}_{ij}$ is a real number and the corresponding Pugh matrix element has only three levels, **+**, **S**, and **-**. The values $\mathbf{C}_{ij}$ are sampled from random variables with distributions $\mathbf{C}_{i1} \sim N(s,1)$ and $\mathbf{C}_{ij} \sim N(0,1), j \neq 1$. Care is required in interpreting the use of random variables here. Random variables enable us to generate a diverse set of concepts, but the values of $\mathbf{C}_{ij}$ are fixed within each trial. The datum concept in the first run has index, *j*=1. The intrinsic merits of the datum concept are selected from a different population than those of all the other concepts. The factor, *s*, represents the relative strength of the datum concept. In our model, larger values are preferred and therefore, if *s*>0, the datum is better than the rest of the concepts on average across the many trials although it can be weak along some criteria in any particular trial.

2. **Model a set of opinions held by a group of experts.** In the model, there are values $\mathbf{CE}_{ijk}$ where $k \in 1 \ldots o$ representing the estimated merit of design concept *j* on criterion *i* as judged by expert *k*. We model the expert opinion as correlated with the intrinsic merits of the

Copyright © 2007 by ASME

design concepts, but differing from expert to expert. This is accomplished by computing the values as $\mathbf{CE}_{ijk} = \mathbf{C}_{ij}\left(1+\varepsilon_{ij}\right)$ with $\varepsilon_{ijk} \sim N(0,\sigma_{ij}^{2})$. Again, these values are related to the Pugh matrix, but not equivalent to it. In particular, there are *o* different expert opinions of each concept's merits along each criterion, yet only one symbol will be entered in the Pugh matrix.

3. **Generate the Pugh Matrix.** Each cell of the Pugh matrix, $\mathbf{M}_{ij}$, corresponds to a design concept *j* and a criterion *i*. The cells are determined as $\mathbf{M}_{ij} =$ **+** if $\mathbf{CE}_{ijk} > \mathbf{CE}_{1jk}$ for all $k \in 1\ldots o$, $\mathbf{M}_{ij} =$ **-** if $\mathbf{CE}_{ijk} < \mathbf{CE}_{1jk}$ for all $k \in 1\ldots o$, $\mathbf{M}_{ij} =$ **S** otherwise. To state the same thing another way, if all experts agree that the concept is better than the datum, then a **+** is entered in that cell. If all experts agree that the concept is worse than the datum, then a **-** is entered. If there is any disagreement among the experts, then an **S** is entered.

4. **Eliminate Concepts Based on the Pugh Matrix.** In actual use of the Pugh Controlled Convergence process, there is no formulaic prescription that automatically leads to the elimination of a concept. In this model, we eliminate any concept that is dominated. In other words, if another concept in the set is better according to **M** along any criterion and is no worse according to **M** along any criterion, then the dominated concept seems to have no unique advantages and will be eliminated.

We simulate the process above to explore the impact of the number of criteria considered and the strength of the datum concept on the ability of the team to eliminate concepts from consideration. Figure 3 presents the simulation results. The strength of the datum was varied from zero to two in ten equal increments (*s*=0, 0.2, 0.4, … 2). Recall that a value *s*=2 implies the datum concept is, on average, two standard deviations above the mean of new concepts generated. The number of criteria was varied across a broad range including 1, 7, and 18. Each point in Fig. 3 arises from 1000 replications of a model with 15 initial concepts and 11 experts.

A principal conclusion we draw from Figure 3 is that the first round of PuCC tends to retain many concepts rather than risk eliminating anything worthwhile. For example, if there are a large number of criteria (such as 18) and *s*=0, then essentially all concepts are retained in the first round. Our interpretation is that, under these conditions, virtually all proposed design concepts possess useful combinations of merits and may prove useful later.

On the other hand, if there is a strong datum, considerable convergence will be attained in the first round and should be relatively insensitive to the precise number of criteria and to variance in expert opinion. If the datum is from a population two standard deviations above the mean of newly proposed concepts, then convergence by more than a factor of two is likely, even with a substantially large number of criteria such as 18. If the datum is one standard deviation above the mean, then only two or three out of 15 concepts will be dominated by the datum if there are 18 criteria. Note that this matches what was observed in Khan and Smith [1989], so we expect *s*=1 provides a reasonably well calibrated model.

When a strong datum is unavailable, and a substantial convergence is desired, the data in Figure 3 suggest that a reasonable option is to focus on just a few strongly discriminating criteria and set all others aside, at least in the first round. Figure 3 reveals that a focus on a small number of criteria will enable considerable convergence even if the datum concept is rather weak. In early stages of design when so many options are in play and there so much uncertainty about most of those options, focus on a small set of criteria can bring clarity to the design process. The team might choose to focus on the few criteria that are most important to the customer or else the ones that define one’s brand. However, it may be better to resist the pressure for convergence in the early runs of the matrix. Subsequent development of our model illustrates that convergence can be accomplished in later rounds.

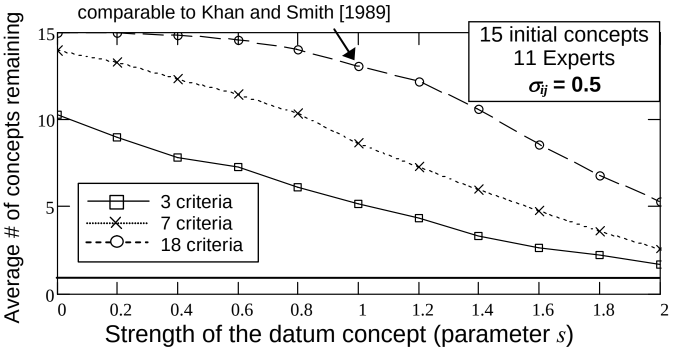

Figure 3: The ability of the first run of the evaluation matrix to eliminate weak concepts.

3.2 A Model of Work between Matrix Runs

We saw in the previous section that the first run of the Pugh matrix eliminates only a modest number of concepts. Each one of the remaining concepts exhibits potential. The process of extracting value from the set of concepts can be incorporated into our model.

When a large number of concepts are in play, some additional decision making is needed to set priorities for further work. This is a principal justification for summary information that is constructed at the bottom of the Pugh matrix. Concepts with a large number of **+** scores and relatively few **-** scores represent good platforms on which to build a serious contender against a strong datum. Concepts with a small number of **+** scores and relatively many **-** scores represent sources of ideas, but probably do not deserve further investment in their own right. The PuCC process does not include any formula for making these decisions. Nevertheless, we propose an algorithm so that we can implement it in our model. The team works in two ways:

Ideation -- Between runs of the matrix, the team can invest time and energy in ideation -- creative work focused by the information revealed in the previous matrix run. This is an important aspect of the engineering work that would normally be conducted between iterations of Pugh's evaluation matrices. Our model of this activity is based on the possibility of forming hybrids of two concepts. Sometimes one can combine different aspects of two or more concepts to form a new concept superior to any of its constituents. We assume that between runs of the matrix, one of the designs in the top 1/3 is selected at random as a basis for a hybrid. Based on the matrix **M** from the last run, a second design is selected that appears most complementary in the sense that it has strengths in just those areas where the chosen concept requires improvements. The hybrid is then formed assuming that, for each criterion *i*, the new value $\mathbf{C}_{ij}$ is the larger of those of two designs being merged. This is an abstract, highly simplified representation of the creative process. In reality, complex technical factors determine which combinations of concepts are feasible and which are not. We want to express in our model the possibility that such hybrids can emerge in response to the evaluation process. We seek to represent this in a reasonably realistic way so that a small number of hybrids that address some, but not all of the observed challenges we observe in experience. This model of ideation, although rough, does enable study of the interplay between creative work, evaluation, and decision making which we believe is critical to drawing an accurate picture of various concept design methods.

Investigation -- Between runs of the matrix, the team can seek improved understanding of the design problem. Because resources are assumed to be constrained, we model investigation of a focused nature guided by the last Pugh matrix. Our model of this activity is that for each concept *j*, if it was in the top 1/3 and it earned an **S** in the previous Pugh matrix on criterion *i*, then for each expert *k* the opinion $\mathbf{CE}_{ijk}$ is refined. In addition, all the concepts receive a refined estimate in the three most influential criteria. The refined estimates are modeled by reducing the parameter σ_{ij} by a factor of two and newly sampling the expert opinion. This is meant to represent the possibility that investigation (including computation, experimentation, interaction with customers, and discussion among the experts) can lead the team to a shared understanding of the issues affecting the decision. In our model, investigation moves the criterion estimates of each expert into better alignment with the objective merits.

Figure 4 presents results from simulations conducted with ideation and/or investigation included as described above in repeated rounds of controlled convergence. The horizontal axis corresponds to the phase of the work with progression in time from left to right. We assumed that the Pugh matrix would be run three times with two periods of work between matrix runs. Each point in Fig. 4 arises from 1000 replications of a model with 15 initial concepts, 11 experts, a moderately strong datum concept (s=1), and moderately large initial variance in expert opinion (σ_{ij} = 0.5). The vertical axis represents the number of concepts under consideration. We ran two cases, one in which a single hybrid concept was formed between matrix runs, and a case with no new concepts generated. For the case including ideation between matrix runs, we plot the median and the 10th and 90th fractiles to give a sense of the variance within the population of trials. For the other case we plot only the median to avoid cluttering the Figure. The convergence observed in a real world case study in Khan and Smith [1989] is also presented for comparison.

A key observation from Figure 4 is that the model with hybrids being generated is generally consistent with the trend in Khan and Smith [1989]. Although there is little convergence in the first round, subsequent work does enable weak concepts to be eliminated to reduce the set of alternatives to a more manageable size. Both the simulation model with hybrid generation and Khan and Smith provided substantially more convergence than the simulation model lacking hybrid generation. Our model of hybrid generation did not attain, in the median, the same level of convergence observed in Khan and Smith, although that high degree of convergence was observed in more than 10% of our trials.

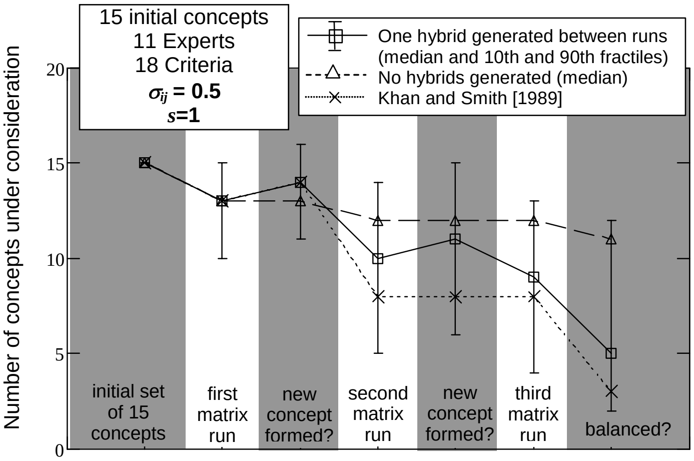

Figure 4: The convergence of PuCC through three iterations with and without new concepts being generated.

It is critical to appreciate the mechanism explaining the connection between divergence and convergence. A hybrid of two complementary designs can often dominate a substantial number of competitors. Visualizing the patterns of strengths and weaknesses in the Pugh matrix seems, based on our experience, to catalyze the creative work needed to generate new concepts that can simplify future decision making. We believe this was the reason that, in Khan and Smith [1989], so many concepts were eliminated after a second run of a Pugh matrix. Our model of hybrid generation included two such hybrid creation events, but still did not match the convergence attained by actual practitioners who reported only one hybrid being generated. We suspect that engineers are better at creating hybrids after running Pugh matrices than we have reflected in our simulations.

Even if we acknowledge the ways that creative work can create dominant concepts, convergence by dominance alone may not suffice for convergence. According to our simulations, if there are many criteria, around half of the total concepts may remain even after three rounds of Pugh matrix runs. However, considerable additional convergence can be made once it is known that additional hybrids will not be formed. As the datum strength increases through PuCC, many designs tend to have one or two positives overwhelmed by a large number of negatives. Although not strictly dominated, poorly balanced designs can be safely eliminated after the last matrix run without sacrificing future opportunities for creative work. Our model suggests that a simple rule based on a 2:1 ratio of **-**:**+** will eliminate about half of the remaining concepts. At this point, either a few designs should be developed in detail, or else recourse might be made to rating and weighting or probabilistic analysis (as in Takai and Ishii [2004]) to converge to a single alternative.

4. COMPARISON OF DECISION MAKING APPROACHES

The previous section shows that the Pugh Controlled Convergence Process, under appropriate conditions, can down-select to a small number of alternatives without resorting to voting, rating, or weighting. But we also need to explore the merits of such an approach compared to alternative procedures. The next sub-section presents an extension of the previous model to incorporate "bottom line" measures of the design outcome. Subsequently that model is used to evaluate methodological alternatives inspired by the design literature such as papers by Hazelrigg [1998], Saari [2004], and Ishii [2004].

4.1 A Model of Profitability

Let us suppose there is real scalar $\mathbf{P}_j$ which represents the overall merits of the j^{th} design concept. It is convenient to think of the **P** vector as standing for profitability of the j^{th} design concept if it were selected and developed.

Central to our model is a quantitative relationship between the criteria $\mathbf{C}_{ij}$ and the value of $\mathbf{P}_i$. We assume that, all other things being equal, a higher rating along one criterion should cause the overall merit to rise. However, we also want to address the issue of "separation of concerns" raised by Saari [2004]. Our model includes the possibility that scoring best across individual criteria does not necessarily imply a design that scores best overall. It is our judgment that this does not happen often in practice, but we include it here to measure its possible impact. To include this possibility and otherwise keep the model as simple as possible, we include only two-factor interactions between pairs of criteria.

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

The sensitivity of the overall merit of any design concept to the i^{th} criterion score is represented by β_i and the interactions among criteria are represented by β_{pq}. By modeling the relationship between criteria and **P** in this way, we are assuming that a full set of criteria uniquely determine the expected outcomes of the design process. In other words, we assume the expected profitability of two designs should be the same for any two concepts that score the same on all criteria.

To instantiate instances of the model in Equation (1), we select the coefficients β from the populations $\beta_i \sim |N(0,1)|$ and $\beta_{pq} \sim N(0,\tau^2)$. The coefficients with a single subscript are non-negative so that the criterion values more naturally correspond with the conventional symbols in the evaluation matrix (e.g., a **+** is meant to indicate a "better" value). The parameter τ represents the relative degree of interactions between criteria. To express the notion that main effects are usually larger than interactions, we suggest τ<<1 in a reasonable model of concept design. Increasing values of τ lead to a situation in which criterion values individually explain only a small fraction of the overall merit. Given the distributions we have chosen, interactions between criteria are equally likely to be synergistic or anti-synergistic.

At this point it is useful to discuss the concept of inter-criterion interactions which are included in our model through coefficients β_{pq}. An improvement in a criterion value such as "manufacturability" should lead to an increase in a measure of overall merit such as expected profit. An improvement in some other criterion, such as "ease of use" should also lead to increase in a measure of overall merit. However, there may be good reasons to believe the effects of two improvements are not simply additive. In extreme cases, anti-synergistic interactions creates a risk of a ranking reversal which was emphasized by Saari [2004]. We do not consider such inter-criteria interactions a major concern in engineering because rank reversals should be rare. However, by including coefficients β_{pq} we allow for the possibility in our model so that we can explore the influence of these effects. If there are only two criteria, then ranking reversal happens only if $\beta_{1,2} < -(\beta_1 + \beta_2)$. Given our model, the probability of this event is $0.5\left[1-(2/\pi)\tan^{-1}\left(\sqrt{2}/\tau\right)\right]$. Therefore we see that the parameter τ enables the modeler to set the probability as desired, but some additional guidance is useful. According to a study by Li, Sudarsanam, and Frey [2006], two-factor interactions in physical experiments are typically about 20% of main effects. If τ is 0.2, the probability of a preference reversal due to an interaction is about 5% per opportunity. In other words, if we choose values of τ typical of physical experiments, then individual criteria and overall merit are consistent in roughly 95% of all instances. With good specification of the design problem, inter-criterion interactions may be smaller than interactions between physical factors, but this is a subject for further research. Those using the model can modify their assumptions in this regard or test the influence of their assumptions by changing the value of τ in the model.

4.2 Profitability of Pugh Controlled Convergence

This section is intended to represent an implementation of Pugh Controlled convergence including a model of profitability. We adapted the model described in Section 3.2 which included three rounds of Pugh matrices to include the profit model presented in Section 4.1. Also, for the purpose of simulation, the final convergence from a handful of options to a single alternative had to be forced somehow. Although Pugh emphasized that this final decision rests with the engineers and not with the matrix, we simply chose the design with the highest difference between the sum of **+** scores and sum of **-** scores in that column of the matrix, a heuristic procedure sometimes called "tallying" [Gigerenzer et al., 1999].

In Figure 5 are plotted the results of these PuCC process simulations. The abscissa represents the average **P** of the selected concept normalized by the maximum value in the initial population of design concepts. The ordinate represents the model parameter σ_{ij} which is can be interpreted as the uncertainty in the criterion scores.

A principal observation from Figure 5 is that the possibility of continuing ideation during evaluation, as was strongly emphasized by Pugh, has a large influence on the outcomes. Figure 5 suggests that even when σ_{ij} is unity meaning uncertainty is as large as the variations within the population, the benefits of continued ideation are larger than the decrements due to uncertainty so that one will attain average **P** values exceeding the maximum value in the initial population. The implication is that no degree of finesse applied to the decision among the fixed set of initial alternatives can compensate for failing to exploit the benefits of additional creative design work.

A second major observation from Figure 5 is that there is very little influence of small degrees of uncertainty on the PuCC process. Figure 5 suggests that even when σ_{ij} is less than 0.5 meaning uncertainty is half as large as the variations within the population, the influence on the outcomes is very nearly zero for all four scenarios we simulated.

Our last major observation from Figure 5 is that the benefits of focused investigation are considerable, especially when uncertainties are large. Figure 5 suggests that even when σ_{ij} is as large as unity and ideation is not used, investigation can remove the majority of the losses due to uncertainty. This is somewhat surprising since investigations in our model are conducted for only about 10% of the criterion/concept pairs. Thus, our model tends to support the notion that Pugh matrices are helpful in locating leverage points for modeling, experimentation, and information sharing among experts.

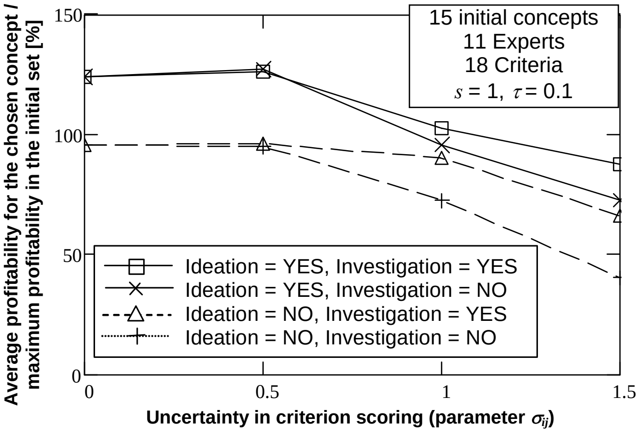

Figure 5: The outcomes after three runs of the Pugh matrix.

4.3 A Model of the Decision Based Design Framework

This section describes an implementation of Hazelirigg's Decision Based Design framework [1998] as applied to the model presented in section 4.1. The scenario simulated is similar to that in section 4.2 except that there is only a single round of evaluation. Also, because the mathematics of decision making as conceptualized by Hazelrigg apply only to individual decision makers, we assume there is just one expert in this model. We formulated a single summary criterion which is that expert's estimate of expected profitability, **PE**. We assume this estimate is related to the true profitability of the concepts but subject to uncertainty so that $\mathbf{PE}_j = \mathbf{P}_j(1+\varepsilon_j)$ with $\varepsilon_j \sim N(0,\sigma_j^2)$. We model the decision maker as risk neutral so that he prefers the highest expected value of profit. Under this assumption, the decision is made simply by picking the largest scalar from among the 15 estimated $\mathbf{PE}_j$ values. We modified the simulation used in section 4.2 to reflect these changes. We set the strength of the datum at a moderate value (s=1) and we ran these simulations for a range of different degrees of uncertainty in expert judgment σ and plotted the **P** value of the selected concept normalized by the maximum value of **P** in the available set of 15 alternatives. The results are depicted in Figure 6 with selected data from Figure 5 also shown for comparison.

The results presented in Figure 6 admit a simple interpretation. When the designer's uncertainty is zero, the profitability is 100% of the potential within the initial set of 15 designs. In other words, if somehow the profitability of the design concepts can be estimated accurately, then choosing the highest estimated profit will obviously maximize the profit. However, the plot shows that as the designer's uncertainty rises, profit attained drops. With a σ of 1.0, only about 75% of the potential profit will be realized on average. Note that given a σ of 1.0, the uncertainty in the evaluation of profitability is roughly as large as the variance among the profitability of the options. This is by no means an upper limit -- the uncertainty involved in estimating profitability at an early stage of the design process might be substantially greater.

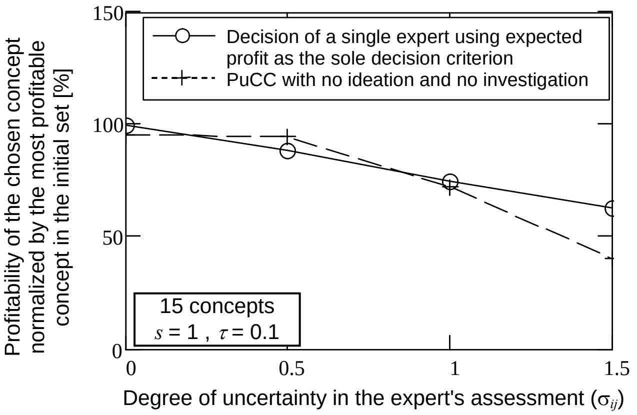

Figure 6: The profit earned based on decisions using a single expert using only estimates of expected profit.

Figure 6 also depicts data from Figure 5 on the performance of PuCC under the worst scenario we considered -- no ideation or investigation allowed. Note that PuCC performs about 5% worse than the DBD approach at zero error. According to our model, the payoff for implementing the DBD framework is that this 5% loss might be avoided. If the resources needed and constraints imposed by DBD detract from ideation or investigation, the net effect of DBD will be negative according to our model.

The two decision procedures plotted in Figure 6 offer very similar performance as a function of the parameter σ with an advantage for the DBD approach as σ rises above unity. However, as discussed in Section 2.5, the research of Smith et al. [1984] show that the reliability of human judgments is better by roughly a factor of two in discrimination tasks on which PuCC is based as compared to magnitude estimation on which the DBD framework is based. If this phenomenon actually applies in engineering design, PuCC might provide better results even in highly uncertain environments.

A preliminary conclusion of this comparative analysis is that internally consistent decision processes can still result in very large losses when uncertainty is high. These losses are due to lack of external correspondence of the decision maker's judgment. By contrast, PuCC may be subject to some potential for internal inconsistencies, but it enables better external correspondence in this model since it involves many experts in the decision and focuses their attention on things that are important to the decision outcome.

4.4 A Model of the Borda Count

This section is intended to represent an implementation of Saari's suggested approach [Saari, 2004] as applied to the model presented here. We consider the possibility of a Borda count over multiple experts using one criterion and also a Borda count over multiple criteria as judged by a single expert.

We modified the simulation from Section 4.3 so that 11 experts were involved in the decision, but only a single criterion, **PE**, was employed and the Borda count was used to combine the information from the experts to choose a single alternative. The results appeared virtually indistinguishable from those from one expert choosing as described in Section 4.3. This confirms that the Borda count generally retains the internal consistency of the DBD framework, but is also similarly sensitive to uncertainty despite involving multiple experts.

We also modified the simulation from Section 4.3 so that 18 criteria were used and the Borda count was used as if the criteria were voting for the winner as Saari [2004] described. The winner of the election was recorded and the **P** value was computed for each trial. We repeated this procedure in 1000 probabilistically independent simulations. This process was repeated for four different values of τ.

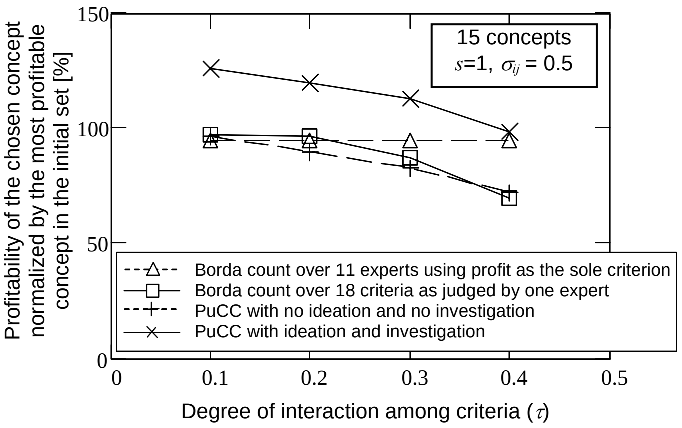

Figure 7: The profit earned as a function of the interactions among criteria.

The results are depicted graphically in Figure 7 with the levels of uncertainty in expert judgment of the criterion values plotted on the abscissa and normalized profitability on the ordinate. A conclusion based on Figure 7 is that, if the criteria are reasonably separable (τ = 0.1 or 0.2), then the Borda count over multiple criteria performs quite similarly to the Borda count over multiple experts based on profit alone. However, if criteria interact strongly (τ = 0.3 or 0.4), then there appears to be a substantial advantage to using just a single summary criterion. We should warn that this assumes that somehow the experts can judge the summary measure (such as expected the profitability) with only as much uncertainty as they would judge the 18 separate criteria. We would venture to say that profitability cannot be estimated as precisely as engineering criteria such as axial stiffness or ease of assembly.

Our preliminary conclusion in this sub-section is that the impact of inter-criterion interactions on PuCC is small for realistic scenarios and is outweighed by what might be lost by diverting attention away from creative work. As noted in Section 3.4, a large meta-analysis of data [Li, Sudarsanam, and Frey, 2006] showed interactions are typically 20% of single factor effects. This would correspond to τ=0.2. But this meta-study data represents interactions among physical factors. We suggest interactions among criteria will be even smaller because: 1) the team of experts are free to define criteria in such a way that they avoid large interactions, and 2) market segmentation tends to limit the degree of differences among concepts considered in PuCC which, a Taylor's series approximation suggests, will encourage a more linear mapping between criteria and overall merit.

4.5 A Model of Rating and Weighting

This section is intended to represent an implementation of one of Takai and Ishii's [2004] proposed variants of Pugh's method as applied to the model presented here.

We modified the simulation from Section 4.3 so that each criterion score is estimated (as a real valued scalar) by a single expert and that linear weighting factors β_i are estimated with the same degree of uncertainty as criterion scores. The ratings and weightings are used to form a score and the concept with the highest score is selected. Note that, given this model, the scores from the rating weighting matrix differ from **P** only because of uncertainty (σ_{ij}) in criterion and weight estimates and because of non-zero criterion interactions β_{pq} (we set τ=0.1). We repeated this procedure in 1000 probabilistically independent simulations. The results of the simulation process are depicted graphically in Figure 8.

A preliminary conclusion based on Figure 8 is that, if uncertainties in criterion scores are the same in both methods (PuCC

Copyright © 2007 by ASME

and rating and weighting), then the outcomes are very similar. But again, research of Smith et al. [1984] show that the reliability of human judgments is improved by about a factor of two in pairwise comparison as opposed to magnitude estimation. If that research applies here, PuCC is preferred to rating and weighting even no new concepts are generated.

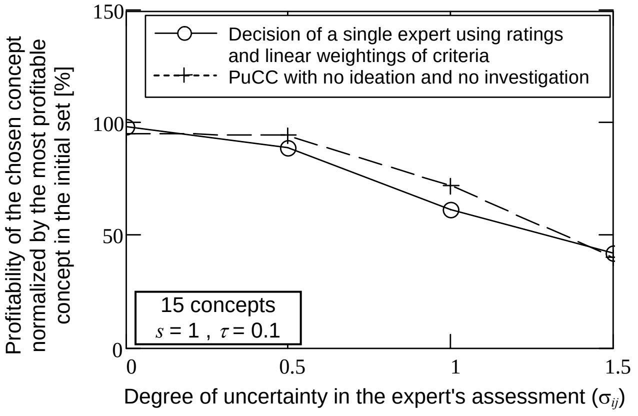

Figure 8: The profit earned based on decisions using a single expert using ratings and linear weightings.

5. CONCLUSIONS

The conclusions drawn from this study must be viewed in light of the degree of model validation conducted thus far. First, as a minimum, the model presented here should have enough common structure with concept design scenarios to enable "surrogative reasoning" [Swoyer, 1991]. In addition, we have set the model parameters in accordance with data from large meta-studies and the results of the model are in reasonable agreement with an industry case study. Although this paper includes substantially more empirical data than the papers criticizing Pugh's method, more validation is required before the model before can serve as a predictive tool.

5.1 Conclusions about Pugh Controlled Convergence

The models presented here support the contention that Pugh Controlled Convergence is an effective method to apply during the concept design phase, mostly because it encourages an interplay of evaluation and ideation. There are risks of internal inconsistencies and distortions as emphasized by Hazelrigg, Saari, and Frannsen (as summarized in Figure 1), but the model suggests that these considerations are outweighed by other issues. Our personal experience with PuCC has been good and the model presented here generally supports this positive evaluation and enables further probing into required assumptions and underlying mechanisms.

One essential conclusion is that ideation and evaluation should proceed in parallel. The PuCC process encourages this. By displaying the merits of design concepts in a clear visual format, Pugh's matrix evaluation procedures prepare the minds of the team for further creative work. What our model suggests is that if just a couple new hybrid concepts emerge from insights into which pairs are complementary, then this consideration alone may trump many other concerns for internal consistency.

Another important conclusion is that uncertainty should not be taken as an immutable facet of design decision making. Engineering design, as it is normally practiced, includes a sequential, iterative process by which uncertainties are reduced through experimentation, investigation, and information sharing among experts. Methods that facilitate this learning process should be strongly encouraged. The model presented here supports the idea that PuCC can help teams target alternative/criterion pairs with high leverage in the decisions they face. In our models, reducing uncertainty in a targeted fashion improved the design outcomes. Similar observations were made by Ward et al. (1995) in studying design at Toyota and Nippondenso where multiple design options are often carried forward, concept selection is deferred, and decisions can be based on more data.

Our model supports the notion that the datum concept is important, especially in the early rounds of PuCC. The practical consequence is that datum selection should not be haphazard. An analysis of the existing competition should be undertaken to identify a concept, perhaps a leader in the marketplace, that can serve as a yardstick for all the others. Our model suggests that a strong datum is likely to simplify decision making and improve the rate of convergence. This conclusion fits well into a broad historical perspective of engineering. Most every successful new design results from evolution of existing successful designs. We conclude that the central role of a strong datum concept in Pugh's method is well aligned with the evolutionary nature of most engineering.

The models presented here also support the notion that Pugh's method encourages greater objectivity in engineering decision-making. In general, people working together on an engineering project should have a substantial agreement on goals (such as whether profit is the dominant objective) and values (such as attitude toward risk). If such agreement is in place, differences of opinion on an engineering team can frequently be settled based on facts. Because of this, engineering decisions may converge as knowledge is shared and evidence is accumulated. In other words, we agree with Scott and Antonsson that there exists a well-defined aggregated order among alternatives and that the availability of this order depends on "time and resources." Pugh's method is intended to facilitate this sort of fact-based convergence. By focusing the team on criteria at an appropriate level of detail, the resulting decisions can be determined more by facts and less by emotional attachments of team members to favorite concepts. Movement in the direction of objectivity, although never realized perfectly, is to be greatly valued.

5.2 Conclusions about Design Theory

The analysis of Pugh Controlled Convergence enables insights into the role of economics and social choice theory as tools for understanding engineering decision making. These theories make assumptions that don't always map well into engineering. For example, Saari's analysis of election procedures assumes each person's stated preference ordering deserves equal consideration. This seems appropriate in a democratic election, but not so appropriate in engineering. Imagine a scenario in which an engineer believes, based on her expertise, that a particular concept is weak and a voting process results in the team selecting that concept. If the dissenting individual based her judgement on facts not known to the others, it provides little comfort that the voting process ensures that her opinion was weighed just as much as every other expert's opinion. We suggest that it would be better to spend time discussing, in concrete engineering terms, her reasons for holding her opinion rather than investing that same time in a process that prevents the distorting effects of Condorcet cycles. The results in Section 4.4 suggest that investigating the reasons for a difference of opinion and exploring new options in light of what is revealed is more productive than using a carefully crafted election procedure to decide the matter.

Franssen [2005] proposes that Arrow's theorem applies fully to multi-criteria decision-making because preferences are "mental concepts neither logically or causally determined by" physical parameters. The implication is that Arrow's stipulated conditions such as Unrestricted Domain and Minimal Liberty imply that preferences must be unrestricted by any demand for objectivity. In personal and political contexts, perhaps people should be unrestricted in this sense, but in an engineering context, it seems inappropriate. If an engineer is faced with a solid body of evidence showing the superiority of one alternative over another, we argue they must either conform to the evidence or else their view is irrelevant to rational engineering decision making. Our model provides a means to explore this contention. The model explicitly includes the concept of objective merits possessed by design concepts (reliability, manufacturability, performance). Such objective merits have a bearing on the bottom line outcomes. During the design process, engineers work to improve these objective merits and also the better characterize them and the way they map to summary measures (like profitability). Good correspondence of expert judgments with facts is essential to good engineering design. In our models, external correspondence breaks down in the limit that σ_{ij} becomes very large. In this case, the expert's estimates are aligned poorly with one another and with facts. Our models suggest that profit earned will drop rapidly as external correspondence breaks down. We conclude it is best to avoid a subjectivist position toward engineering decision-making.

5.3 Suggestion for Future Research

There has been much discussion of rationality in engineering design and almost all of the emphasis in this discussion has been on internal consistency. Sen [1993] has argued that internal consistency demands "cannot be assessed without seeing them in the context of some external correspondence, that it, some demand originating outside the choice function itself." We suggest this represents a great opportunity for research concerning external correspondence and its role in engineering decision-making. It seems to us that a theory of engineering design, recognizing the important role of decision making, must have something to say about not only how data is processed by an individual, but also how data is gathered via interaction with the real word.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 0448972.

REFERENCES

Box, G. E. P., and N. R. Draper, 1987, *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, Inc., Hoboken, NJ.

Czerlinski, J., G. Gigerenzer, G., and D. G. Goldstein, 1999, "How good are simple heuristics?," In G. Gigerenzer, P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart* (pp. 97-118), Oxford University Press, New York.

Constable, G. and B. Somerville, 2003, *A Century of Innovation: Twenty Engineering Achievements That Transformed Our Lives*, National Academies Press, Washington, DC.

Diederich, A., 1997, "Dynamic Stochastic Models for Decision Making Under Time Constraints," *Journal of Mathematical Psychology* **41**: 260-274.

Franssen, M., 2005, "Arrow’s theorem, multi-criteria decision problems and multi-attribute preferences in engineering design," *Research in Engineering Design* **16** (2005), 42-56.

Frey, D. D., and C. Dym, 2006, “Validation of Design Methods: Lessons from Medicine”, *Research in Engineering Design* **17**(1)45-57.

Gigerenzer, G., P. M. Todd, & the ABC Research Group (Eds.), *Simple heuristics that make us smart*, Oxford University Press, New York.

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.

Hazelrigg, G.A., 1998, “A Framework for Decision-Based Engineering Design,” *ASME Journal of Mechanical Design*, **120**, pp. 653-658.

Hazelrigg, G.A., 1999, “An Axiomatic Framework for Engineering Design,” *ASME Journal of Mechanical Design*, 121, pp. 342-347.

Johnson, J. G. and J. R. Busemeyer, 2005, "A Dynamic, Stochastic, Computational Model of Preference Reversal Phenomena," *Psychological Review* 112 (4) 841-861.

Khan, M., and D. G. Smith, 1989, "Overcoming Conceptual Barriers -- By Systematic Design," *Proceedings of the Institute of Mechanical Engineers ICED*, Harrogate, UK.

Li, X., N. Sudarsanam, and D. D. Frey, 2006, “Regularities in Data from Factorial Experiments,” *Complexity* 11(5)32-45.

Pahl, G., and W. Beitz, 1984, *Engineering Design: A Systematic Approach*, Springer-Verlag, Berlin.

Pugh, S., 1981, "Concept Selection: A Method That Works" Proceedings of the International Conference on Engineering Design ICED, Rome, Italy.

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

Pugh. S., and D. Smith, 1976, "The Dangers of Design Methodology," First European Design Research Conference, Portsmouth, UK.

Saari, D. G., and K. K. Sieberg, 2006, “Are Partwise Comparisons Reliable?,” *Research in Engineering Design* 15: 62-71.

Salonen, M. and M. Perttula, 2005, "Utilization of Concept Selection Methods -- A survey of Finnish Industry," *ASME Design Engineering Technical Conferences*, Long Beach, CA.

Scott, M. J. and E. K. Antonsson, 1999, "Arrow's Theorem and Engineering Design Decision Making," *Research in Engineering Design* **11**(4):218-228.

Sen, Amartya, 1993, "Internal Consistency of Choice," *Econometrica* **61**(3):495-521.

Sen, Amartya, 1998, "The Possibility of Social Choice," *Nobel Prize Lecture*, Trinity College, Cambridge, UK.

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

Solow, R. M., 1957, "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics* **39**(3): 312-320.

Swoyer, Chris, 1991, "Structurql Representations and Surrogative Reasoning," *Synthese* 87:393-415.

Takai, S., and K. Ishii, 2004, "Modifying Pugh's Design Concept Evaluation Methods," *DETC2004-57512, ASME Design Engineering Technical Conferences*, Salt Lake City, UT.

von Neumann, J., and O. Morgenstern, 1953, *The Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ.

Ward, A, J. K. Liker, J. J. Christiano, and D. K. Sobek, 1995, "The Second Toyota Paradox: How Delaying Decisions Can Make Better Cars Faster," Sloan Management Review 36(3):43-61.

Copyright © 2007 by ASME