AC 2009-425: COHERENCE AND CORRESPONDENCE IN ENGINEERING DESIGN EVALUATIONS

Jonathan Evans, MIT
Jonathan R. Evans earned his Bachelor's degree in Mechanical Engineering from the University of Utah in 2007. He has spent time designing devices ranging from nuclear waste processing equipment to special effects technology. He is currently working on his Master's degree at the Massachusetts Institute of Technology in Mechanical Engineering. His research focuses on the application of decision-making cognitive methods to the engineering design process.

Konstantinos Katsikopoulos, Max Planck Institute for Human Development
Konstantinos V. Katsikopoulos received his Ph.D. in industrial engineering and operations research from the University of Massachusetts Amherst in 1999. Since then, he has held visiting faculty positions at the Naval Postgraduate School and the Massachusetts Institute of Technology (the latter visit was partly funded by a German Science Foundation Fellowship for Young Researchers). He is currently a senior research scientist at the Center for Adaptive Behavior and Cognition of the Max Planck Institute for Human Development. He works on the modeling of decisions that people (laypeople and experts) make or should make with limited time, information, and computation.

Chad Foster, MIT
Chad Foster earned a BS degree in Engineering at the Colorado School of Mines in 1998. He then went on to earn a MS degree in Mechanical Engineering from the University of Texas at Austin in 2000. After spending five years working as an Engineer at Cummins Inc., he returned to school and is currently pursuing a PhD at the Massachusetts Institute of Technology. His primary research interests are around early-phase robust design and improving experimental techniques.

Coherence and Correspondence in Engineering Design Evaluations

Abstract
Much of the formal education in engineering design focuses on the teaching of analytical methods. Analysis allows one to make coherent statements about the performance of designs. There are situations in design and design education, however, where it is appropriate to use intuition: A focus on intuition sometimes allows one to make statements that correspond well with the real performance of designs. Here we studied such a situation.

Students in a design course competition were asked to make quick evaluations of the performance of other students' designs. The surveys also contained either analysis-inducing questions or intuition-inducing questions. We found that the students put in the intuitive, correspondence-based mode, evaluated design performance more accurately. While, given this finding, the correspondence mode is more appropriate in this situation, we found a lack of consensus among design instructors and practitioners about which mode would be more effective.

In sum, our results indicate that the engineering curriculum should include methods of correspondence-based, intuitive decision-making and that instructors should be sufficiently aware of these methods to help students identify situations where they should be employed.

Introduction
The majority of formal education in engineering focuses on analytical problems and analytical solutions to those problems. There are many decision situations in the design process, including educational design, which could benefit from intuitive or methodologies.

During a student design course, students were presented with a decision situation evaluating the performance of alternative designs. This situation was created to favor the intuitive methodology, and it was shown through student surveys that they were more accurate when in the intuitive mode versus the analytic mode. On the other hand, there was a lack of consensus among instructors about the performance difference when asked which mode would be more effective.

Engineering curriculum should include aspects of intuition-based decisions and help students identify situations where they are more effective rather than solely relying on analytical decision-making methods.

Background
Engineers must make decisions repeatedly throughout their careers. They are confronted with multiple design, material, or manufacturing alternatives and must decide which alternative presents the best design option. The thought processes used for these decisions are not very well understood. Research in the field of decision-making has produced two metrics for good decision-making: coherence and correspondence. Research by Kenneth Hammond[5,6,7] has shown that these two metrics are not exclusive, as was once thought, but are actually complementary. These two metrics, correspondence and coherence, describe the underlying goal of the decision making process while intuition and analysis describe the process.

Correspondence and Intuition
In correspondence-aiming decision-making, the goal is empirical accuracy. It relies on the natural ability of humans to take information from multiple sources and evaluate it "almost instantaneously"[5]. Many of these sources are fallible (may be incorrect), like visual estimations about distance and speed. The brain automatically takes multiple, fallible sources and combines them into a decision about the situation[4]. This is the type of judgment used when trying to determine if one has sufficient time to cross the street before an oncoming car, or when judging the aesthetics of an item. Judgments based on correspondence can happen quickly and almost subconsciously, but they can also produce errors. The goal of correspondence isn't logic; it is empirical accuracy[11]. This means that these decisions are usually based on intuition, experience or ad-hoc testing.

Coherence and Analysis
Coherence-aiming decision-making relies heavily on logic and analysis. Its goal is logical consistency. This is the type of judgment used when doing mathematics, evaluating the logic of a politician's argument, or examining a set of data for inconsistencies[11]. This type of decision-making is not as natural as correspondence. Humans are not innately born with the ability to rationalize decisions. That is one reason it is emphasized in school. People must be taught to think logically and to back up decisions with rational arguments[11].

Hypothesis
These two metrics of decision-making have been applied to many different fields. Hammond used much of his research studying public policy decision-making and economics[5]. Kathleen Mosier applied these same principles to research about aircraft piloting[11]. Current decision frameworks in all these fields rely heavily on both coherence and correspondence. Little has been done, however, to apply these metrics to the field of engineering design. Emphasis in engineering education, as is the case in almost all areas of formal education, centers on analytical decision-making (which leads to coherent choices). The classes often present the students with analytical or mathematical assignments. These work at developing the logical, analytical skills needed to solve such problems and justify those solutions. The hypothesis presented here is that, in certain situations, correspondence decisions, relying heavily on intuition, can lead to more accurate results and that an understanding of those situations should be emphasized in engineering education.

Research Methods/Setup
A second year mechanical engineering class provides students at the Massachusetts Institute of Technology (MIT) with a chance to learn about and implement design skills. The course is centered around a design contest where the students build robots to accomplish certain tasks. The contest for the spring 2008 course involved building robots that could pick up rubber balls, foam cylinders and toy beavers. Each item had to be placed in a different container. The students were individually responsible for designing, manufacturing, testing, and controlling their robot for the competition. The complete rules for the contest can be found on the course website <http://pergatory.mit.edu/2.007/>.

Student Survey

For this research, the students participating in the class were given a chance to fill out a survey on the day of the contest. This survey asked them to compare several robots built by the other students. Ten students volunteered to have their robots evaluated. These robots were displayed on a table to allow the evaluators to examine them. The students were asked to make comparisons in a pair-wise manner, making four sets of comparisons in total. (i.e. Student A might have been asked to compare robots 1 and 2, robots 5 and 6, robots 7 and 6, and robots 9 and 10.) They were given a small prize for participating.

The survey was designed to account for performance differences depending on the table location of the robots under comparison as well as a difference between the priming questions. The initial design was created to minimize the biases problems because of location effects, either different locations on the table or the order of the survey questions.[14]. If the robots were too far apart to be visible at once the participant may be influenced by only having one of the comparisons available, and has to rely on memory for the other robot. This could lead to a biased decision by only recalling the most frequent or unusual characteristics of the robot. The second problem could be anchoring with the first robot in the comparison and interpreting comparisons with favoritism towards that initial robot. In addition to keeping the robots close the order of robot comparison was also randomized. There was no evidence in the data that these effects were evident, and the probability that these variables could yield this data was very low (p=0.4, 0.2). The performance of the robots was unknown when the survey was designed or these effects could have been accounted for more specifically.

Each robot was pair-wise compared to the two robots on either side giving a possible four comparisons for each robot. The order of the comparison was randomized and each person did not review the same comparison twice. The utility of pair-wise design comparisons has been shown to have many potential problems[13,8],however there have been results that highlight the strengths of paired comparison in early design evaluations[3]. The major objection with pair-wise comparisons is that it is impossible to determine an accurate rank given a number of pair-wise votes. The objective in this case was slightly different; the goal was to predict the outcome from pair-wise competitions, not to rank overall winners. The competition in general may have problems with the order of matches, and may not lead to the best robot winning (or other Condorcet type problems[2]). In this case the assumption is that there will not be a Condorcet problem and that there is a single rank order of the robots. In addition, there has only been one tie in the 40 year history of the competition.

For each comparison, the students were asked three questions. Two of the questions concerned different aspects of the robots, such as power consumption or number of parts. The last question for each comparison asked the students to decide which robot would win in a head-to-head contest between the two. The students were allowed to look at the robots and examine them, but not touch or otherwise interfere with the robots.

The purpose of the first two questions (the priming questions) was to get the students to make decisions in either a correspondence or coherence mode. Half of the priming questions were coherence-based and asked questions that encouraged the students to think analytically. The other priming questions were correspondence-based and encouraged the students to think more

intuitively. The last question asked to determine the best performing robot, and differed slightly between the correspondence and coherence modes.   Each student was asked to make two comparisons with coherence questions and two with correspondence questions (in random order). These questions, along with a sample survey, are found in Figures 1 and 2.

Figure 1: Student Survey Questions

Coherence Questions
1. Which robot will use more battery power?
2. Which robot has more parts?
3. Which robot has done a better job of properly supporting moving parts?
4. Which robot weighs more? (no touching robots)
5. *Last Question always asked:* What would each robot score in a head-to-head contest? (circle the winner)

Correspondence Questions
1. Which robot is faster?
2. Which robot is more robust?
3. Which robot is more elegant?
4. Which robot is more complex?
5. *Last Question always asked:* Which robot would win in a head-to-head contest?

Figure 2: Sample Student Survey

Sample Survey

For each set of machines listed below, quickly compare the designs. Remember that this is a quick comparison; don't spend more than a minute on each set of machines. Use this sheet of paper also as a scratch sheet if you want to do any calculations/counting regarding each of the machines.

MACHINE A   MACHINE B

_____   _____

     A   1. Which robot weighs more? (no touching the robot)   **B**

     A   2. Which robot will use more battery power?   **B**

3. What would each robot score in a head-to-head contest? (circle the winner)

A			B	
Food Balls	_____		Food Balls	_____
Noodles	_____		Noodles	_____
Beavers	_____		Beavers	_____

MACHINE A   MACHINE B

_____   _____

     A   1. Which robot has more parts?   **B**

     A   2. Which robot is has done a better job of properly supporting moving parts?   **B**

3. What would each robot score in a head-to-head contest? (circle the winner)

A			B	
Food Balls	_____		Food Balls	_____
Noodles	_____		Noodles	_____
Beavers	_____		Beavers	_____

MACHINE A   MACHINE B

_____   _____

     A   1. Which robot is more robust?   **B**

     A   2. Which robot is faster?   **B**

3. Which robot would win in a head-to-head contest?
     A   **B**

MACHINE A   MACHINE B

_____   _____

     A   1. Which robot is more complex?   **B**

     A   2. Which robot is more elegant?   **B**

3. Which robot would win in a head-to-head contest?
     A   **B**

The setup of the experiment was designed to favor intuition and correspondence. According to research done by Hammond et al.,[6] several factors can influence the prevalence and effectiveness of intuitive (heavily correspondent) or analytical (heavily coherent) cognition. The task surface characteristics (how the information is presented) and the task depth characteristics (what information is asked of the subject) both influence the type of cognition used.

Their research showed that surface characteristics that presented information in a simultaneous, graphical way (such as pictures and physical models) favored intuition. Conversely, information presented numerically and sequentially (such as formulas and equations) favored analysis. The type of questions that asked for information about the robot for which there was no available formula also favored intuition. Those questions that involved counting or calculating favored analysis.

The students' main source of information was the robots themselves. They were only allowed to look at the robots, providing them with multiple fallible visual cues. They were given no data about these robots and did not have much time to perform calculations. Each of these factors biased the situation towards intuition[6]. The students were given the option to perform calculations on the back of the survey, but none did.

The evaluation of the effectiveness for the type of priming question was based on the quality of their guesses compared with the actual performance. In a competition with this number of robots and a heat-based-layout each robot does not compete with every other robot. A performance ranking was created by experts (local designers and professors) to allow the comparison of every robot. Additionally, the actual competition performance was not included because of the high variance attributed to wiring glitches (motors) or driving errors. The goal was to judge the design potential of the robot and not the driving ability or other complications. The performance was based on a ranking that was given by a consensus the experts.

A couple of ranking differences existed between the expert ranking and the scores achieved during the competition, there was one robot that was ranked quite low but had a very high average score in the actual competition. When the actual scores were used there were a number of challenges including robot degradation, large score variation, and opponent influence. One contestant scored 282 points in one round and 0 the next, or another contestant jumped from 100 points to 1248 points. This variation could be contributed to the scoring strategy during that round, and this was not consistent between rounds. There were a number of strategies that were difficult but, if successful, achieved high scores.

The other scoring variation was due to the influence of the other robot. Some contestants used a blocking strategy, where they would score first and then prevent their opponent from scoring, thus lowering both scores. Thus the raw scores from the robots were too variant to achieve significance on any of the variables of interest. The final analysis used an expert ranking that is similar to how the class grades are assigned, and represents a very good performance estimate. There was approximately 98% Pearson correlation between the expert ranking and the competition performance. The important aspect for the hypothesis here is that there was good correlation between the priming questions and the accuracy in predicting the winner.

Instructor Survey

After giving the survey to the students, several of the instructors for the course, as well as several design professors, were asked to predict the student performance on this estimation. In particular given the set-up, which set of priming questions, correspondence or coherence, would lead to better results? This group was selected because of their design experience and familiarity with the course/contest. The instructors were presented with information about the students' survey and the questions asked. A copy of this questionnaire is shown in Figure 3.

Figure 3: Sample Instructor Survey

Survey

Below are two sets of questions that we asked 60 students who participated in the course design contest. Each student made four different comparisons of 10 robots side-by-side without touching the robots. In two comparisons they were asked questions from Set A, and in the other two they were asked questions from Set B. The students were then asked which robot would perform better in a contest (the last question in each set). They were not given a time limit on the comparisons, and there were both high and low performing robots. In the end the accuracy of their judgments were compared with the actual robot performance rankings and the question set.

As a knowledgeable designer, we would like your input about which set of questions would be more predictive of robot performance. That is, which set of questions would lead a majority of students to a more accurate assessment of robot performance in the competition?

Question Set A:
1. Which robot will use more battery power?
2. Which robot has more parts?
3. Which robot has done a better job of properly supporting moving parts?
4. Which robot weighs more? (no touching robots)
5. What would each robot score in a head-to-head contest? (circle the winner)

Question Set B:
1. Which robot is faster?
2. Which robot is more robust?
3. Which robot is more elegant?
4. Which robot is more complex?
5. Which robot would win in a head-to-head contest?

More accurate set of questions: _____A   _____B

Reasons:

The professors gave their responses either orally during interviews or by filling out the survey sheet. Their responses were then compared with the results from the students.

Results
Of the students enrolled in the design course, 49 filled out surveys. The students' responses to the last questions (those asking about the performance of the robots) were compared to the performance ranking developed. The only statistically significant factor in the evaluation performance was the type of priming questions used (coherence or correspondence). The type of the priming questions was significant ($p=0.03$) in a logistic regression. Correspondence-inducing priming questions led to more accurate evaluations by the students. Given the short timeframe for the analysis and that only visual cues were used, the better performance of the correspondence priming questions fits within the framework developed by Hammond. The situational influence to the predictive ability of students through priming questions identifies a challenge and an area for improvement in design decisions. Certain situations exist where performance is improved by utilizing correspondence, but identifying those situations is difficult for students as well as instructors.

On the instructor questionnaire there was no consensus: half of the instructors thought correspondence would be better and half coherence. It would have required 53 instructor surveys to conclude if they were truly split rather than 70% for one method or the other. While there were insufficient instructors to establish that level of confidence, we can say that there is not a 90% consensus among the instructors that either mode would be superior. And while this is not conclusive, it does offer insight into the current attitudes among engineering design instructors. In many of the discussions with the instructors they pointed out deficiencies with the survey that prevented coherence based thought such as: "The student should be given the questions earlier…", "they should be given the weight/part count for the robots." Even with this insight they did not reach the conclusion that correspondence would yield better predictions.

The majority of the course focused on coherence based thinking and the application of those methods to the designs. There was no formal instruction in the course to help students identify situations where intuition based thinking would be more appropriate.

Correspondence Instruction
This paper centers on recognizing the opportunity for correspondence based judgments without going into methods to acquire or improve that ability. Within this study we assume the correspondence ability of participants based on previous experiences and activities. Other research is ongoing that looks at activities to gain experiences quickly and improve judgments in these situations. One promising area is by Klein[10], around recognition primed decision making. These methods assist in re-experiencing situations so you can benefit from the situations of others. This research has not focused on design judgments specifically but on other time-sensitive situations, specifically fire-fighting. Investigating specific ways to enhance correspondence judgment in a design context is a promising area for future research.

Other options could include introducing more concept-based testing, such as the Force Concept Inventory[9], into classrooms. These types of tests tend to be less analytical and encourage a more intuitive understanding of engineering principles. Current research is underway studying what impact students' conceptual understanding of engineering principles has on engineering design ability and success.

Conclusions

As mentioned above, the student survey experiment was intentionally biased towards intuition and correspondence. As would be expected in this situation, the students performed better when asked correspondence priming questions. The instructors, however, did not consistently agree about which type of questions would be more effective. This demonstrates the challenge in identifying the limits of a methodology when applied in an inappropriate situation. Much of the formal education in engineering focuses on analytical problems and analytical solutions to those problems. There are many situations in the design process, including educational design, which could benefit from an awareness of the potential of intuitive or correspondence based methodologies.

With the continued evolution of design instruction, the value and appropriateness of some non-analytical decision methodologies should be included. This could help to prevent inappropriate application of decision methods, not only for the case here, where intuitive insight is advantageous, but also for situations where decisions should be analytical. In the course the instructors have frequently seen students avoid analysis for sizing shafts, specifying motors, or fastening parts. In these areas calculations are available, appropriate and accurate.

In other areas, however, analytical methods can require too much time or unavailable information[1]. In situations demanding rapid decisions with external cues as the primary information source, intuitive answers can sometimes be more efficient and effective. The ability to differentiate between these situations, and, if necessary, even combine both the intuition and analytical methodologies is needed for real-world design. Putting an accurate scope and weight on analysis and accepting that some decisions may be better based on correspondence could benefit the design curriculum and better prepare future engineers.

Special Thanks

Works Cited

1. Clausing, Don P., Konstantinos V. Katsikopoulos. "Rationality in Systems Engineering: Beyond Calculation or Political Action." *Systems Engineering.* Vol. 11, No. 4, Winter 2008.
2. Condorcet M (1785) E´ssai sur l'application de l'analyse a` la probabilite´ des de´cisionsrendues a` la pluralite´ des voix. Paris
3. Frey, Daniel D.; Paulien M. Herder; Ype Wijnia; Eswaran Sbrahmanian; Konstantinos Katsikopoulos; and Don P. Clausing. "An Evaluation of Pugh Controlled Convergence Method" *Proceedings of ASME DETC: Design Engineering Technical Conference* 2007
4. Hamm, Robert M. "Moment-By-Moment Variation in Experts' Analytic and Intuitive Cognitive Activity." *IEEE Transactions on Systems, Man, and Cybernetics.* Vol. 18, No. 5, September/October 1988.
5. Hammond, Kenneth R. *Human Judgment and Social Policy: Irreducible Uncertainty, Inevitable Error, Unavoidable Injustice.* Oxford University Press, NY, 1996.
6. Hammond, Kenneth R., Robert M. Hamm, Janet Grassia, and Tamra Pearson. "Direct Comparison of the Efficacy of Intuitive and Analytical Cognition in Expert Judgment." *IEEE Transactions on Systems, Man, and Cybernetics.* Vol. SMC-17, No. 5, September/October 1987.

7. Hammond, Kenneth R. *Beyond Rationality: The Search for Wisdom in a Troubled Time.* Oxford University Press, NY, 2007.

8. Hazelrigg, George A. "The Implications of Arrow's Impossibility Theorem on Approaches to Optimal Engineering Design" *Journal of Mechanical Design.* 1996

9. Hestenes, David, Malcolm Wells, and Gregg Swackhamer. "Force Concept Inventory." *The Physics Teacher.* Vol. 30, March 1992.

10. Klein, Gary A., "Sources of Power: How People Make Decisions", MIT Press, Cambridge, Mass, 1998

11. Mosier, Kathleen. "Achieving Coherence: Meeting New Cognitive Demands in Technological Systems". *Adaptive Perspectives on Human-Technology Interaction.* Oxford University Press, NY, 2006.

12. Mosier, Kathleen L. "Cognition in the Automated Cockpit: A Coherence Perspective." *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting – 2001.*

13. Saari, D. G. and K.K. Sieberg. "Are Partwise Comparisons Reliable?" *Research in Engineering Design* Volume 15, 2004.

14. Tversky, Amos and Daniel Kahneman. "Judgment under Uncertainty: Heuristics and Biases" *Science* Volume 185, 1974.