The Importance of Training for Interactive Trade Space Exploration: A Study of Novice and Expert Users[1]

David Wolf
Graduate Research Assistant

Jennifer Hyland
NSF REU (Research Experiences for Undergraduates) Student

Timothy W. Simpson[2]
Professor
ASME Fellow

Department of Mechanical and Nuclear Engineering,
The Pennsylvania State University,
University Park, PA 16802
e-mail: tws8@psu.edu

Xiaolong (Luke) Zhang
Assistant Professor
Department of Information Sciences and Technology,
The Pennsylvania State University,
University Park, PA 16802

Thanks to recent advances in computing power and speed, engineers can now generate a wealth of data on demand to support design decision-making. These advances have enabled new approaches to search multidimensional trade spaces through interactive data visualization and exploration. In this paper, we investigate the effectiveness and efficiency of interactive trade space exploration strategies by conducting human subject experiments with novice and expert users. A single objective, constrained design optimization problem involving the sizing of an engine combustion chamber is used for this study. Effectiveness is measured by comparing the best feasible design obtained by each user, and efficiency is assessed based on the percentage of feasible designs generated by each user. Results indicate that novices who watch a 5-min training video before the experiment obtain results that are not significantly different from those obtained by expert users, and both groups are statistically better than the novices without the training video in terms of effectiveness and efficiency. Frequency and ordering of the visualization and exploration tools are also compared to understand the differences in each group's search strategy. The implications of the results are discussed along with future work.
[DOI: 10.1115/1.3615685]

Keywords: multidimensional data visualization, design optimization, assessment, user training

1 Introduction

Engineers routinely use computer-based simulation and analysis models to support design decision-making [1], particularly during the parametric and detailed stages of design when optimization tools can be employed. Optimization tools provide one means to explore multidimensional trade spaces to find the design solution that maximizes (or minimizes) one (or more) objective while satisfying relevant constraints [2]. Recent advances in computing power and speed have enabled new interactive approaches to search trade spaces using multidimensional data visualization and exploration tools [3,4]. Such approaches allow designers to "steer" the optimization process while searching for the best (or Pareto optimal) design(s) [5,6], and recent studies have shown significant gains in the computational efficiency by putting designers back "in-the-loop" during the trade space exploration process [7].

To support interactive trade space exploration, researchers at Penn State University and the Applied Research Laboratory (ARL) have been developing the ARL trade space visualizer (ATSV) since the early 2000s [8,9]. ATSV has evolved into a platform for conducting research into human-computer interactions (HCIs) by allowing us to study how designers use multidimensional data visualization tools to display and navigate complex trade spaces to find design solutions [10]. What has become increasingly apparent in these studies is the importance of user training, not only in terms of using the software and its capabilities but also in terms of interpreting visual displays that involve different representations of multidimensional data. The issue of training is not particular to the capabilities in our software (see the comparison offered in Ref. [10]), yet it provides a unique opportunity for us to study it in the context of engineering design.

In this paper, we investigate the effectiveness and efficiency of interactive trade space exploration strategies by conducting human subject experiments with novice and expert users solving a single objective, constrained design problem. Our distinction between novices and experts derives from their experience with the visualization and exploration tools available in our software and not with the problem domain. The capabilities of our software are summarized in Sec. 3 following a review of related literature in Sec. 2. The experimental setup, test problem, and user trials are described in Sec. 4, and the results and their implications are discussed in Sec. 5. Section 6 provides closing remarks and avenues for future work.

2 Review of Related Work

Since our software (i.e., ATSV) uses data visualization as the main form of user feedback from the system, it is important to understand the differences between novices and experts with respect to using data visualization tools. Seo and Shneiderman [11] find that interactive exploration of multidimensional datasets can be challenging because it is difficult to see patterns in more than three dimensions. Klein [12] states that expertise is based on a person's ability to recognize and match patterns. The ability to perceive patterns and then to match patterns to actions in decision-making is

[1]An earlier version of this paper appeared in the 2009 ASME Design Engineering Technical Conferences, ASME, San Diego, CA, Paper No. DETC2009/DAC-87294.
[2]Corresponding author.
Contributed by the Simulation & Visualization Committee of ASME for publication in the Journal of Computing and Information Science in Engineering. Manuscript received March 5, 2010; final manuscript received April 11, 2011; published online September 2, 2011. Assoc. Editor: Ian Grosse.

built up through experience and practice [13]. From this we can gather that novices may not have yet developed the pattern recognition ability of expert users and therefore may struggle with higher dimensional data visualization. Viewing data in more than three dimensions also makes it harder to discover relationships, outliers, clusters, and gaps in the data [11]. Pattern detection is important especially for developing user-centered methodologies for interactive trade space exploration because humans are capable of learning from patterns and using this knowledge to improve their performance in a manner unattainable by current algorithms [14]. Additionally, Petre and Green [15] find that both perceptual and interpretive readership skill for graphical representations must be learned. Thus there is a clear difference between the way novices and experts utilize data visualization tools [16] as less experienced users are unable to interpret graphical cues that may be helpful. Among the strategy differences between novices and experts, Petre and Green [15] state that novice users tend to confuse visibility with relevance; conversely, experts are able to match patterns and disregard irrelevant information.

Many aspects of the interactive trade space exploration process are akin to naturalistic decision-making, which attempts to describe how decisions are actually being made in the field. Naturalistic decision-making also involves various goals and subgoals that are likely to change as new information is received and priorities change [13,17]. Engineering design problems often contain conflicting goals and time constraints that characterize the circumstances for naturalistic decision-making [18]. Recognition-primed decisions are said to occur in naturalistic decision-making [12], where experienced decision-makers are able to spend more resources assessing a situation rather than assessing different courses of action. Experienced decision-makers do not use their resources to generate a list of possible decisions before making a decision; rather, they draw from previous experience to accept and reject decisions one at a time. This provides the user with improved situational awareness and enables the decision-maker to work better under time constraints by being continually prepared to initiate an action [13]. In this way naturalistic decision-making is influenced by the expertise of the decision-maker. Studies have shown that experts place emphasis on situational assessment, while novices emphasize deciding the course of action [13].

Since expert users have developed the ability to identify the appropriate path to a solution, they process information in a nongoal specific manner [19]. Conversely, novices tend to work backward from the solution, which does not promote knowledge toward nongoal specific problem solving. Since experts use generalities to work toward the solution of a problem, they are better able to use their knowledge to solve varying problem types that trade space exploration presents. Another distinction between novices and experts is the relative frequency that they use specific processes [20]. Novices tend to use a passive strategy of collecting data and seeing what happens, whereas an expert's ability to reason results in a much more varied mix of decision-making processes. Moreover, according to cognitive models of novices and experts [21,22], novices need to retrieve knowledge from declarative memory frequently to perform tasks while experts do not. This should translate into experts using a wider range of tools during interactive trade space exploration because they are able to better see the big picture, whereas novices may become confused by the process and number of data elements [12]. This is partially due to the fact that novices treat every piece of information as an independent unit [23] while experts use "chunking" to treat several distinct items of information as a single unit. This allows experts to track more relevant information and have better situational awareness than novice decision-makers [20]. Chunking can also help experts develop suitable mental models to reduce errors that commonly occur in tasks that require situational awareness [24].

Finally, user expertise affects what a decision-maker needs from a decision-aiding program such as ATSV. Expert users desire rapid response times, brief and nondistracting feedback, as well as the ability to carry out actions with a limited number of commands [25]. Novice users, on the other hand, require informative feedback about task accomplishment as well as effective support methods toward task completion such as instructions, dialog boxes, and online help. In order for a novice to carry out tasks successfully, a limited number of actions should be required [25]. Specifically, Shneiderman [25] suggests that users be allowed to control the density of information feedback that a system provides. Similarly, they should be allowed to control the density of displays, as expert users prefer displays that are more densely packed than novices. In light of this, a summary of ATSVs multidimensional data visualization tools and visual steering capabilities for decision-aiding is given next.

3 Visualization and Exploration Tools

3.1 Multidimensional Data Visualization Tools. ATSV is a Java-based application developed to support trade space exploration research [4]. Thus, ATSV is capable of visualizing multidimensional trade spaces using glyph, 1-D and 2-D histograms, 2-D scatter, scatter matrix, parallel coordinate plots, linked views [26], and brushing [27]. Figure 1(*a*) shows a glyph plot that can display eight-dimensional information using the spatial position of an icon to represent three variables of a data point; an additional five variables can be represented by the glyph's size, color, orientation, transparency, and text overlay (only size and color are used in this example). Multiple histogram plots can be displayed within a single window as shown in Fig. 1(*b*). Parallel coordinate plots, shown in Fig. 1(*c*), represent designs using polylines [28] that intersect parallel axes representing data dimensions. A scatter matrix [see Fig. 1(*d*)] is used to view multiple combinations of 2-D scatter plots.

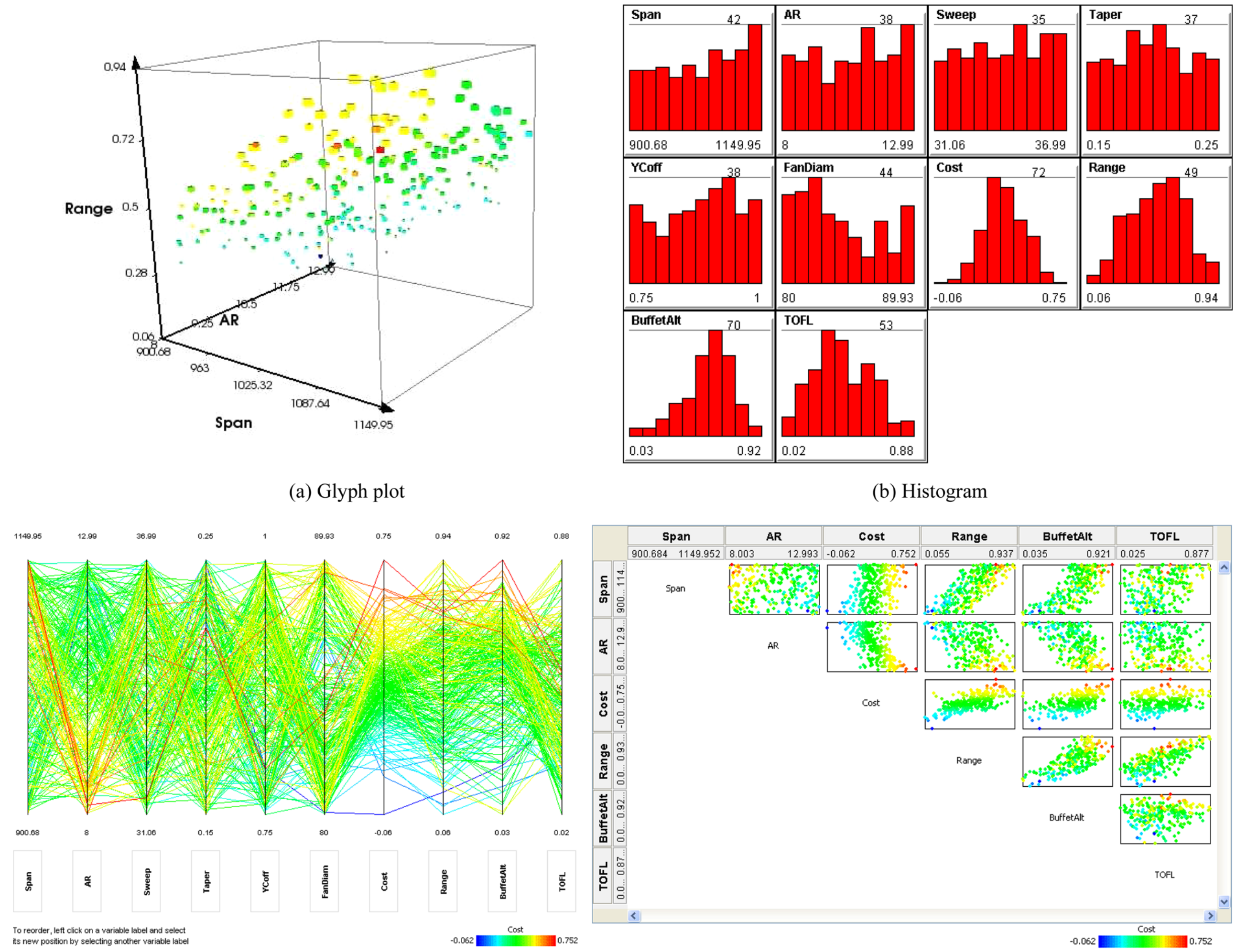
3.2 Visual Samplers and Exploration Tools. ATSV offers a variety of exploration tools, as introduced in Ref. [4], to allow designers to visually guide the generation of new designs within the trade space. These tools currently available include: (1) basic sampler, (2) point sampler, (3) attractor sampler, (4) preference sampler, and (5) Pareto sampler. A brief description of each follows.

Basic samplers are used to populate the trade space and are typically invoked if there is no initial data available. The user specifies the number of samples to be generated and the bounds of the multidimensional hypercube. Monte Carlo sampling [29] is used to randomly sample the inputs and execute the simulation model, and the corresponding output is stored in a database. The bounds of the design variables can be reduced at any point to bias the samples in a given region.

Point samplers allow the user to manually sample the design space by moving slider bars for each input variable. As such, this sampler allows designers to perform one-factor-at-a-time studies of the simulation model instead of random sampling. After moving a slider bar, the simulation model is executed at the design point specified by all of the slider bar settings.

The *attractor sampler* is used to generate new sample points near a user-specified location in the trade space. The attractor is specified in the ATSV interface with a graphical icon that identifies an n-dimensional point in the trade space, and then new sample points are generated near the attractor—or as close as they can get to the attractor. Unbeknownst to the user, the attractor generates new points using the Differential Evolution (DE) algorithm [30], which assess the fitness of each new sample based on the normalized Euclidean distance to the attractor. As the population evolves in DE, the samples get closer and closer to the attractor. An example is shown in Fig. 2 where the user places attractor_1 to try and generate aircraft wing designs that have a low cost and high range (see Ref. [10] for the problem description).

Preference-based samplers allow users to populate the trade space in regions that perform well with respect to a user-defined



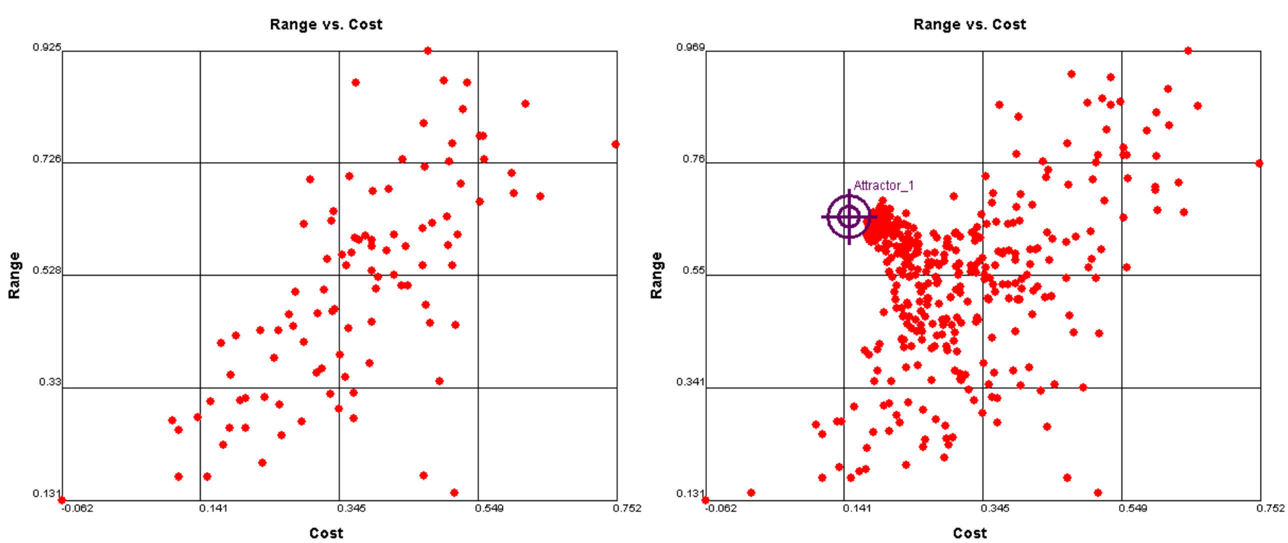
(a) Glyph plot

(b) Histogram

(c) Parallel coordinates

(d) Scatter matrix

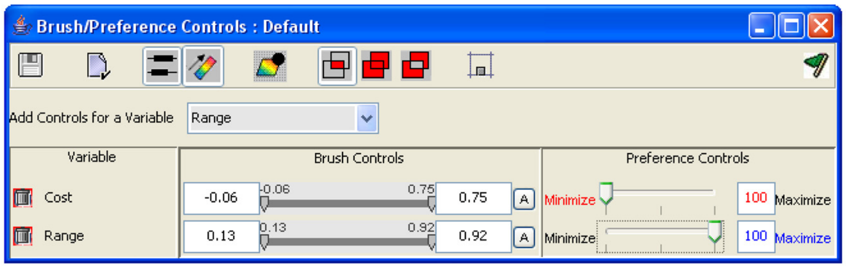
Fig. 1 Multidimensional data visualization examples



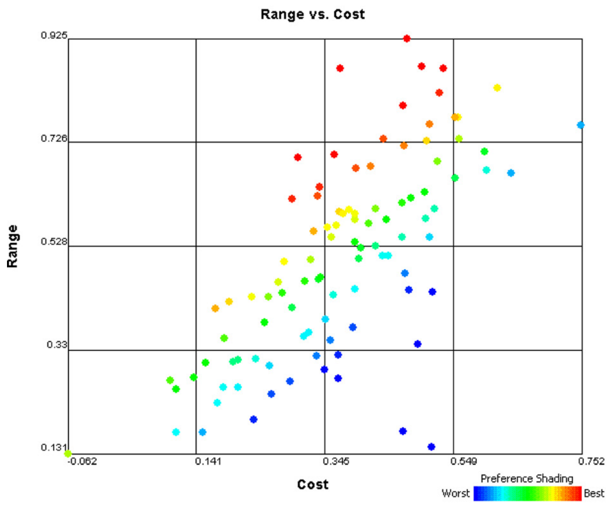
(a) 100 initial samples

(b) New samples generated near attractor

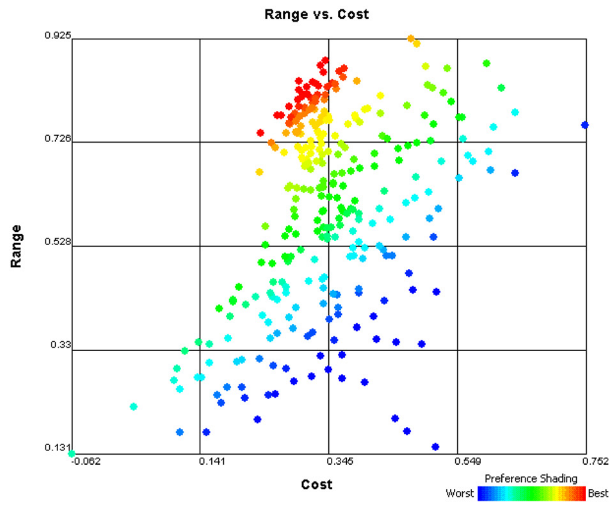
Fig. 2 Example of attractor sampler



(a) Brush settings indicating user preference structure



(b) Initial samples colored by preference

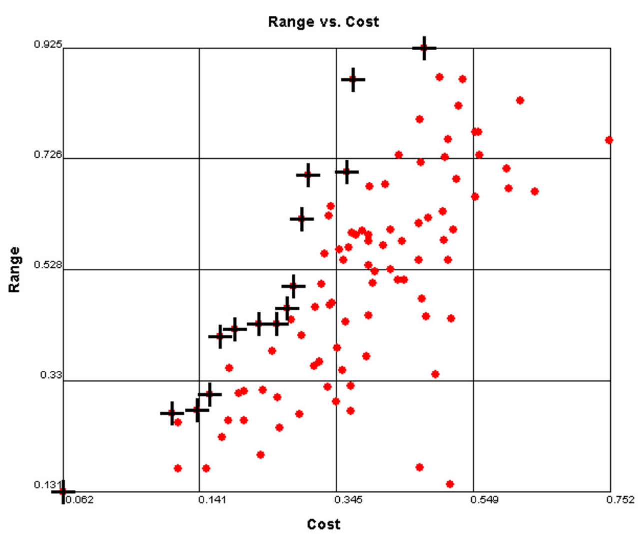


(c) New samples generated in preferred direction

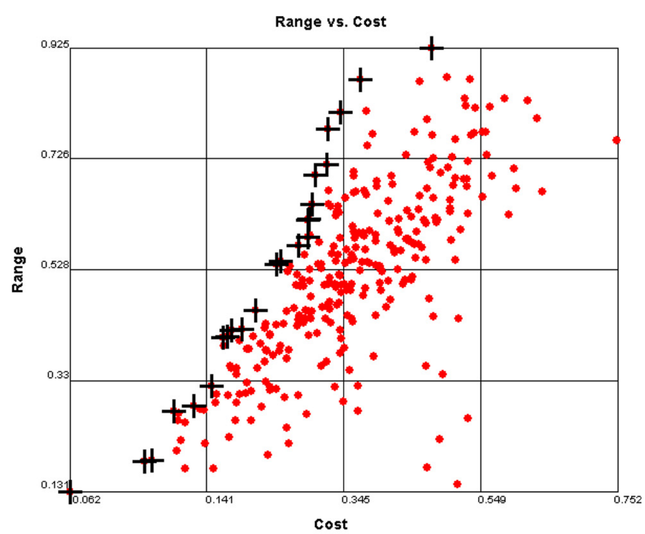
Fig. 3 Example of preference-based sampler

preference function. New sample points are generated by the DE algorithm, but the fitness of each sample is defined by the user's preference structure instead of the Euclidean distance. An example is shown in Fig. 3. Using ATSV's brushing and preference controls, the user specifies a desire to minimize cost and maximize range of an aircraft with equal weighting [see Fig. 3(*a*)]. Figure 3(*b*) shows the initial samples shaded based on this preference, and Fig. 3(*c*) shows the new samples that concentrate in the direction of preference.

Pareto samplers are used to search for the Pareto frontier once the user has defined his/her preferences for the objectives. The DE algorithm is used for this sampling but modified to solve multi-objective problems [31]. An example of this sampler is shown in Fig. 4. Using the same preference as before (i.e., minimize cost and maximize range), Fig. 4(*a*) shows the Pareto points in the initial samples; Fig. 4(*b*) shows the Pareto frontier after executing seven additional generations of the DE with a population size of 25 points per generation.



(a) Initial samples



(b) New samples generated along Pareto frontier

Fig. 4 Example of Pareto sampler (Pareto points denoted by +)

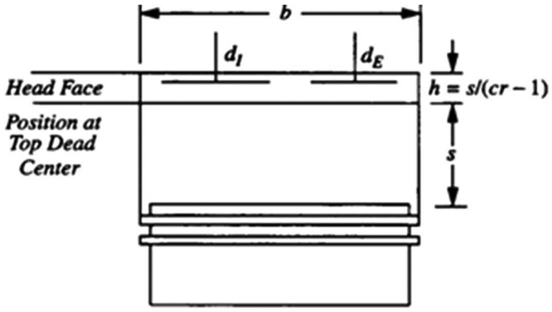


Fig. 5 Definition of combustion chamber design variables [34]

Table 1 Inputs and bounds for combustion chamber

Design Variable	Full Name	Lower Bound	Upper Bound
b	cylinder bore, mm	70	90
d_I	intake valve diameter, mm	25	50
d_E	exhaust valve diameter, mm	25	50
c_r	compression ratio	6	12
w	revolutions per minute at peak power, ÷1000	5	12

Using ATSV's multidimensional data visualization and exploration tools can add complexity to the decision-making process. Individual data visualization tools afford different analysis tasks and serve different purposes. Designers need to choose the appropriate tool(s) and settings at each step. Designers also need to know how to combine the individual tools to support decision-making. For instance, the output of one (e.g., identification of a "hole" in a glyph plot) may serve as the input to another (e.g., specification of an attractor to explore this region). Properly sequencing the visualization tools can greatly enhance the decision-making process; unfortunately, such knowledge is not intuitive or obvious to users and training is required. Section 4 describes the experimental setup used to investigate the use of these visualization and exploration tools to support decision-making during trade space exploration.

4 Experimental Setup and User Trials

The development of ATSV and our testing methods have been influenced by work with practitioners [10,32], and many well-established usability inspection methods for HCI exist [33]. For the purposes of this analysis, we use a combination of empirical and formal testing with human subjects to test their interactive search strategies while evaluating user performance. This section describes the test problem, the performance measures, and the participants involved in the study.

4.1 Test Problem. The test problem used in this study is the engine combustion chamber design model [34] depicted in Fig. 5. It is a single objective, constrained optimization problem with five input (design) variables. The upper and lower bounds of the five design variables are listed in Table 1. The original analyses can be found in Ref. [34], and we use the formulation in Ref. [35], which divides the problem into equations pertinent to engine geometry and thermodynamics.

The objective in this problem is to maximize the specific power of the combustion chamber, or in this case to minimize the negative specific power (NSP). NSP is a function of b, d_I, c_r, and w; it is indirectly affected by d_E, which impacts the sizing of the chamber. In the study, participants are only asked to optimize the combustion chamber using the geometry analyses that are summarized in the brush/preference control settings in Fig. 6. As seen in the figure, the Geometry subsystem has six constraints that limit the stroke, bore wall thickness, engine height, valve structure, and valve diameter (minimum and maximum). The constraints are formulated so that values less than or equal to zero yield a feasible design.

4.2 Performance Measures. Two performance measures were used to assess user effectiveness and efficiency. The first was the best feasible design obtained by the user, i.e., the design with the lowest NSP value that satisfied all of the constraints. The best feasible design indicates users' effectiveness with the visualization and exploration tools. The nature of the objective function allows designs to have a very large range of NSP values, from over +1000 to around −60 (as seen in the brush controls in Fig. 6). Given that only negative NSP values are acceptable, it is easy to see how large positive values of NSP can skew the dataset given the larger range of positive NSP values. Thus, the Modified Thompson Tau [36] technique was used to identify and discard outliers.

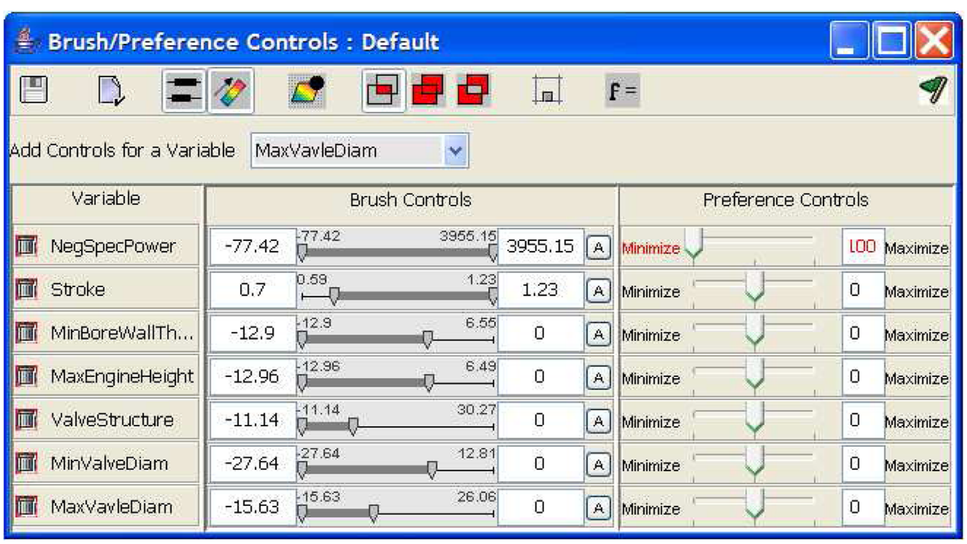


Fig. 6 Brush/preference control settings for the geometry subsystem

The second measure was the percentage of feasible designs generated during a single trial, which provides a measure of the users' efficiency with the tools. Since relatively few designs were generated by some of the users in the study, this metric is more informative than the total number of designs that were generated. Thus, decision-makers who have a higher percentage of feasible designs are considered more efficient in their use of the visualization and exploration tools since they wasted less computational resources on infeasible designs.

In addition to these two measures, we also look at the correlation between the two measures to see if there are any trends. We also study the frequency that each visualization and exploration tool is used as well as the transitions between specific tools. Investigating these aspects of the results provides insight into the users' strategies for trade space exploration.

4.3 Description of Participants and User Trials. Three sets of data were obtained for our experiment: (1) novices without a training video, (2) novices with a training video, and (3) expert users. Our distinction between novices and experts derives from their experience with the visualization and exploration tools (i.e., ATSV) and not with the problem domain (i.e., the combustion chamber). They all had nearly the same level of domain expertise, in that no one had any experience solving the combustion chamber problem prior to this experiment. The novice users consisted of juniors, seniors, and first-year graduate students in mechanical and industrial engineering. These students were recruited randomly through email and in-class announcements, and they were not compensated for their participation in the study. Novices were recruited for the without video group while the training videos were being developed; then a second set of novices were recruited for the training video group. Meanwhile, expert users were research associates from the Applied Research Laboratory that had 3–6 yr of experience developing and working with ATSV. They were recruited through direct contact given the limited number of them available. While many of the experts had had advanced graduate studies, knowledge about multi-objective optimization and trade studies was not relevant given the single objective nature of the combustion chamber problem. In total, 60 participants gave us consent to use their data: (i) 27 novice users without video, (ii) 27 novice users with video, and (iii) six expert users.

For the user trials, both sets of novice users received a fifteen minute overview of ATSV, which described its visualization tools and exploration capabilities. The second set of novices saw an additional 5 min training video after the overview before the experiment. The training video was designed to demonstrate effective visualization and exploration strategies based on an earlier pilot study [37] applied to a different test problem. Expert users skipped the overview and the training video given their familiarity with ATSV.

After reading an overview of the combustion chamber problem and providing consent to use their data, users were given 10 min to solve the problem using ATSV. When time expired, users were instructed to submit their log file that recorded all of their actions in ATSV, and no personal information was recorded in the log file as users were told before providing consent. From this log file, we were able to quantify the frequency that each visualization and exploration tool was used by each participant and how many designs they generated during the analysis. The log file also listed all of the feasible design points (i.e., designs that satisfied all of the constraints) that were generated by each user, from which the total number of designs, the number of feasible designs, and the best feasible design were extracted for our analysis.

5 Analysis and Discussion of Results

5.1 Analysis of NSP Values (Effectiveness). The distributions of the best feasible designs (i.e., lowest NSP values) for each user group are plotted in Fig. 7. Outlier points in the novice groups were removed in accordance with the discussion in Sec. 4.2 to avoid skewing the datasets. Even though a majority of the novices performed with similar results, including the outliers in the data would have skewed the mean NSP value for all novices from −42.31 (with no outliers) to 48.01. Removing these outliers actually benefits the novice groups by making their performance, on average, better for analysis. As a result, statistically significant differences are all the more meaningful.

Figure 8 plots the average best feasible NSP value for each group after removing outliers. Comparisons between groups was performed using two-sample t-tests (two-tailed un-paired sample

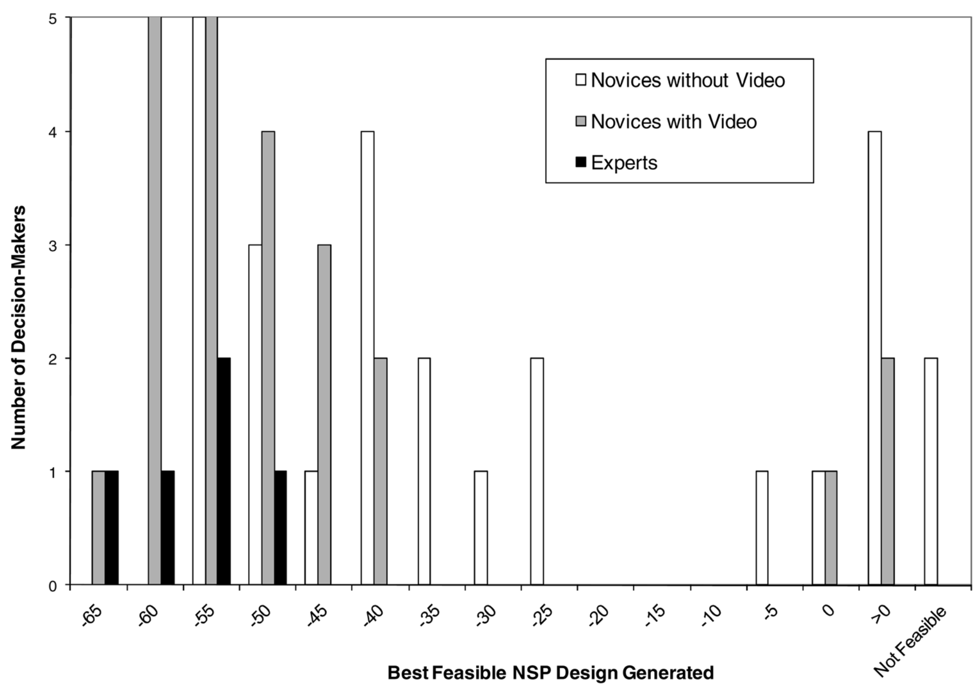


Fig. 7 Distribution of best feasible designs for novice and expert groups

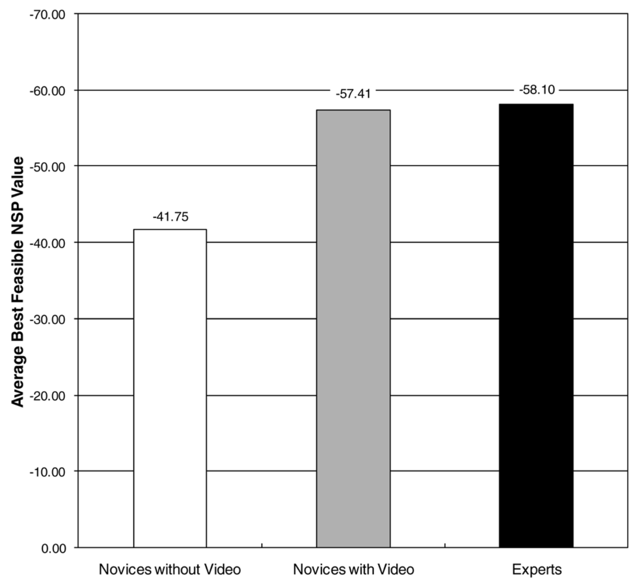


Fig. 8 Average NSP values for novice and expert groups

testing [38]), and p-values less than 0.05 indicate a statistically significant difference between datasets. As evident from the figure, there was a statistically significant difference between the novices without video and the novices with video ($p < 0.001$) and experts ($p < 0.001$); however, there was not a statistically significant difference between the performance of novices with video and the experts ($p = 0.405$). This result is encouraging and surprising. It is encouraging to see that even a minimal amount of training (5 min video) can have a large impact on the performance of novice users, and it is surprising that their performance is indistinguishable from the experts given the large disparity with the novices without the video. In fact, we randomly generated ten sets of 833 points (the average number of points that the novices without video generated) and found that there was no statistically significant difference between the results from the random samples and the novice users without video. This is a discouraging result, but it reinforces the importance of user training.

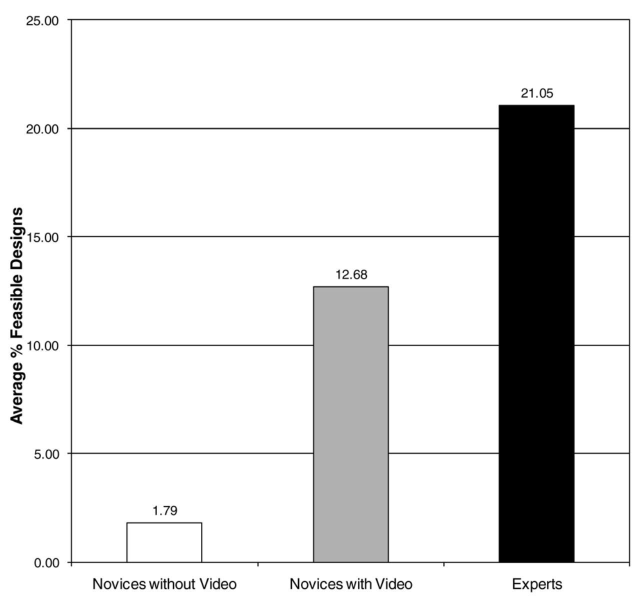


Fig. 9 Comparison based on average percentage of feasible designs

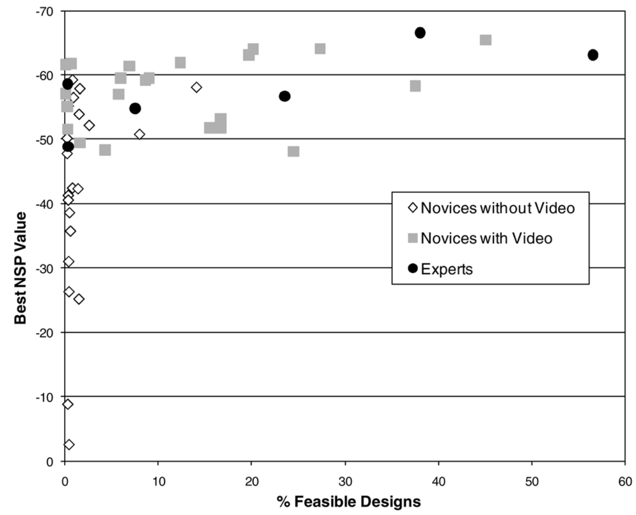


Fig. 10 Scatter plot of percentage of feasible designs and NSP values

5.2 Analysis of Percentage of Feasible Designs (Efficiency). Figure 9 shows the average percentage of feasible designs for the three user groups. As before, the experts and novices with video were able to produce significantly higher percentages of feasible designs ($p < 0.001$) compared to the novices without the video. This suggests that the novices with video, like the experts, are much more efficient when exploring the trade space and focusing their designs in an area of interest. While it looks like the experts are considerably more efficient at finding feasible designs (21.05% versus 12.68%, on average), there is not a statistically significant difference between the experts and novices with video ($p = 0.211$). Figure 10 sheds light on this.

In Fig. 10, we see a positive correlation between the percentage of feasible designs and lower NSP values for the experts and novices with video, but we also observe wide variation in the experts' ability to find feasible designs in the trade space. Upon further inspection of the data, we find that the problem lies with the overall number of designs generated. In one extreme, an expert generated 13,707 designs, of which only 37 (=0.27%) were feasible; meanwhile, another expert generated only 4075 designs, of which 2306 (=56.59%) were feasible. The novices with video were equally scattered in their ability to identify feasible designs in the trade space, and therefore, the two groups are not statistically different. To gain more insight into how these three user groups explored the trade space, we delve into specific tool usage next.

5.3 Analysis of Trade Space Exploration Strategies. In addition to the two performance measures, we also examined the frequency that each visualization and exploration tool was used by each group as well as the transitions between tools to gain insight into different users' exploration strategies. The usage statics are discussed next. Section 5.3.2 discusses the state transition activity diagrams [39] to gain further insight into the specific strategies and procedural knowledge used by the different groups.

5.3.1 Usage of Visualization and Exploration Tools. Figure 11 shows the frequency that each visualization and exploration tool was used by each user group. We can immediately note the limited range of visualization tools that the novices used without the training video employed. Almost every novice user in this category relied on the 2-D scatter plots as their primary method for visualizing the trade space. From the log files it could be seen that students used many variations of 2-D scatter plots and sometimes used multiple scatter plots simultaneously. This result suggests that novice users without the training video are more comfortable

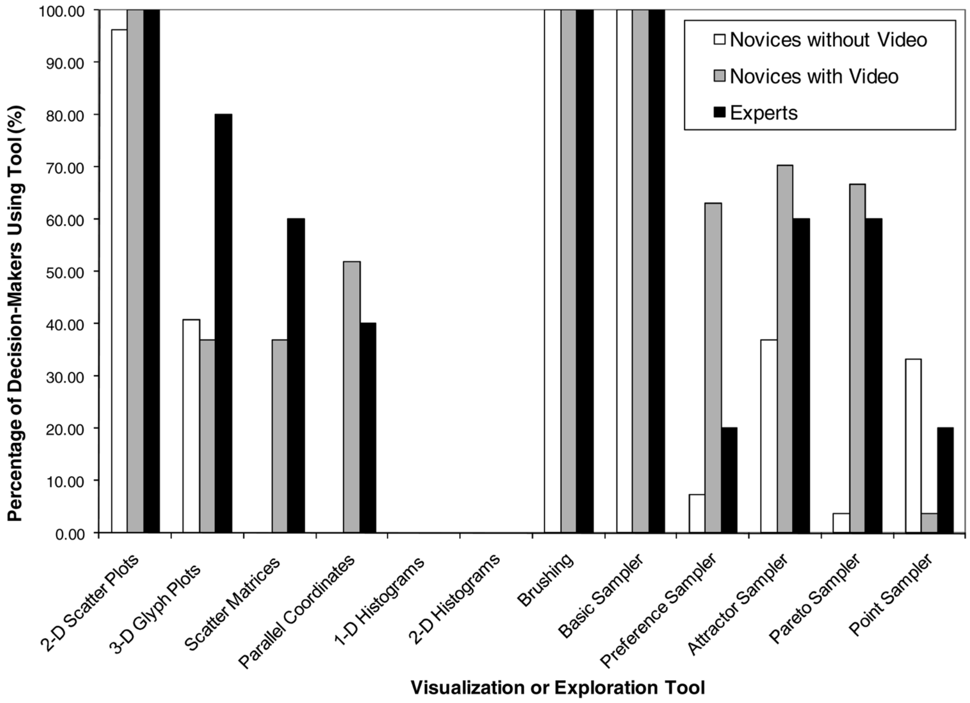


Fig. 11 Percentage of users utilizing each visualization and exploration tool

using lower-dimension visualization tools as fewer than half of them used 3-D glyph plots.

Figure 11 also shows that the novices without the training video never used any of the higher dimensional visualization tools (e.g., scatter matrices, parallel coordinates, 1-D and 2-D histograms), even though they were introduced to these tools. This may indicate a tendency for the novices to use the visualization tools with which they are most familiar and not try new tools when learning how to use ATSV. However, since none of the experts used histogram plots either, this may indicate that they were not appropriate for this particular design problem. Unlike the novices without the training video, expert users employed a wide variety of visualization tools both individually and as a group. One of the goals of the training video was to guide the novice users to use these higher dimensional visualization tools to help them better understand and visualize the trade space. The video sought to encourage users to employ a wider variety of visualization tools to help them explore the trade space while also helping them understand when each visualization tool was situationally appropriate. This goal seems to have been accomplished as the novices who viewed the training video took advantage of a wider range of visualization tools similar to the experts. As seen in Fig. 11, the usage of scatter and glyph plots remained constant across novice groups, but there was a large increase in the percentage of novice decision-makers who used the scatter matrix and parallel coordinate plots after viewing the training video.

The results for the exploration (i.e., sampling) tools in Fig. 11 show similar results to the visualization tools. Independent of the training video, all of the novice users were able to use brushing to specify their constraints, and every novice user also used the basic sampler at least once to sample the trade space. Since it is commonplace to start a trade space exploration problem with a basic sampler run, we expected this sampling method to be used by everyone. What is interesting though is the number of designs the novices without the video generated through random sampling as opposed to other methods. Most of these novices used the basic sampler as their primary sampling method. They were most likely just trying to sample until they found feasible designs, which indicates that they were goal-oriented but did not have the problem solving strategy necessary to best reach their objective. Learning from expert users to understand effective nongoal-oriented problem solving strategy is desirable.

Although the novices were introduced to all of the sampling methods, only one-third of novices without the video used the attractor or point sampler. Very few novices without the video utilized the preference or Pareto sampler. The expert users employed a wider range of sampling methods similar to their use of a wide range of visualization tools. As demonstrated by the expert users, all of the exploration tools can be useful during trade space exploration, and it was important to capture this in the training video to encourage novice users to employ these tools when appropriate. With the training video, the novices used a much wider variety of exploration tools. As seen in Fig. 11, these novices used the preference sampler, attractor sampler, and Pareto sampler with much higher frequency compared to the novices without the training video. Using the point sampler less may have been a consequence of the time constraint and the increased usage of the other samplers. All of the samplers can be very useful in trade space exploration; therefore, it should be a priority to guide and encourage users to use these tools when appropriate.

5.3.2 State Transition Activity Diagrams. Figure 12 illustrates the state transition activity diagrams [39] for each user group. In these figures, the circles represent the different exploration tools and the squares represent the different visualization tools used by the users. The size of the circles is proportional to the percentage of designs generated using that exploration tool across the total population of that respective dataset. Similarly, the size of the squares is proportional to the number of times that visualization tool was utilized. The arrows between the squares and circles represent transitions between specific tools (e.g., the novices without video tended to switch frequently between the basic sampler and the scatter plot). The weight (thickness) of each arrow is proportional to the number of times a transition between those tools was performed. Arrows do not represent the transitions that were not repeated or that occurred infrequently. The information used to create these diagrams was obtained from the log files

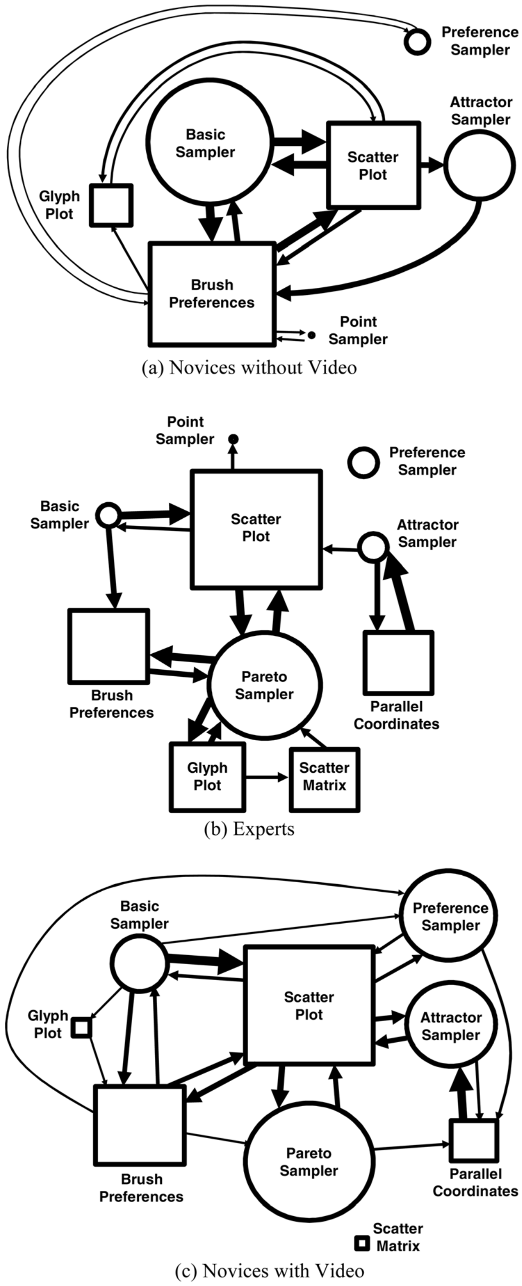


Fig. 12 State transition activity diagrams

by manually coding transitions between tools and recording the number of designs generated with each sampler.

Figure 12(*a*) represents the average state transitions for the novices without video. From this diagram we can see that they relied heavily on the basic sampler and the brushing/preferences tool. The scatter plot was used as the primary method of visualizing the data with the glyph plot being used only occasionally. These novices did not utilize the scatter matrix, parallel coordinates, or histogram plots as noted earlier. They used the attractor sampler the most frequently; however, nearly twice as many designs were generated with the basic sampler. The point sampler and Pareto sampler were used to generate a relatively smaller number of designs in comparison. In terms of the transitions, there were many transitions between the basic sampler and the scatter plot as well as between the basic sampler and brush references. There are also quite a few transitions between the brush preferences window and the scatter plot. Most of these novices' transitions were between three tools (basic sampler, scatter plot, and brush preferences) showing that they are primarily utilizing the simplest tools in ATSV. These novices also often used the attractor within the scatter plot indicating a preference for lower-dimensional tools.

Figure 12(*b*) shows the average state transitions for the expert users. Similar to the novices without video [see Fig. 12(*a*)], the experts utilized the scatter plots with the greatest frequency; they also frequently used parallel coordinates, scatter matrices, and the glyph plots when appropriate. One of the notable differences between the novices without video and the experts is that the experts used the basic sampler much less frequently. Not captured in Fig. 12(*b*) is the fact that experts primarily used the basic sampler to initially populate the trade space and then only used the more advanced samplers thereafter. The experts generated most of their designs using the Pareto sampler, which is in sharp contrast to the novices without training. The attractor and preference samplers were moderately used, while the point sampler was used to generate relatively few designs.

From Fig. 12(*b*) the most common expert transitions can also be seen. In particular, parallel coordinates to attractor sampler was the most frequent transition, showing that the experts preferred to use multidimensional attractors combined with parallel coordinates to visualize the trade space. The strongest two-way transition was between the Pareto sampler and the scatter plot. Many experts also transitioned between the basic sampler and the scatter plots. This shows that after initially populating the trade space, the experts liked to view the designs that were generated before specifying their preferences and constraints.

Figure 12(*c*) represents the state activity transitions for the novices with the training video. Similar to the novices without video, the novices with video primarily used the scatter plot for visualization; these novices also used parallel coordinates frequently and the scatter matrix occasionally. The greatest difference is that these novices relied significantly less on the basic sampler and instead utilized a wider variety of sampling methods. The Pareto sampler, which was not used by the novices without the training video, was used to generate the most designs. The preference sampler was used to generate a large number of designs, most likely in an attempt to generate some feasible solutions as suggested in the training video. The attractor was also used to generate a larger percentage of designs; additionally, parallel coordinates replaced the scatter plot as the preferred visualization tool for the attractor sampler. The training video demonstrated the use of higher dimensional attractors using parallel coordinates, and it appears that many of the novices chose to replicate this strategy. The novices with the training video also appear to be doing a better job of utilizing the visualization tools to guide their design decision-making. They often visualized the initial population of data before inputting the constraints and preferences, which is similar to the experts' approach. Their ability to utilize the visualization tools to guide their design generation is also supported by the fact that there were not a lot of transitions between different samplers, but rather, they are referencing the visualizations both before and after new design generation. Figure 12(*c*) also shows an even distribution of actions similar to that of the experts in Fig. 12(*b*). This suggests that the novices are gaining some procedural decision-making knowledge through the training video and are not simply relying on repeatedly using the simplest tools as with the untrained novices in Fig. 12(*a*).

5.4 Implications of the Results. The results suggest that without the training video, novice users were able to utilize the visualization and exploration tools with limited success to find regions of the trade space that had lots of feasible designs and good NSP values. In fact, the novices without the video were not able to significantly improve their objective in the combustion chamber problem compared to random sampling. Interestingly, untrained novices using higher dimensional visualization tools (e.g., 3D glyph plots) actually performed worse than the whole population of novices [37]. *This suggests that without proper training, novices are unable to use higher dimensional plots effectively to find regions with feasible designs*. The utility of this training was demonstrated through the improved performance among novices who viewed the training video.

Engineering design problems are very diverse, and the most appropriate visualization and exploration tools to use will vary from problem to problem. Understanding in what way individual tools, or combinations of tools, can be used to gain a better understanding of the problem is something that is not easily trained and is perhaps best learned through experience. The training protocols developed in this analysis substantially improved the novices' average performance by teaching them effective design decision-making methodology and making them more comparable to the experts. Since novice users initially do not have a large range of situational experience to draw upon, training protocols helped teach novices how to find patterns in the data and to filter out irrelevant information. Not captured in the results was the fact that novices tended to misuse the visualizations they selected by not viewing the dimensions of greater importance. In this way, they exhibited the novice tendency of confusing visibility with relevance [14]. With the training video, the novices were better able to visualize the dimensions of higher importance, namely, the objective, through parallel coordinates, or by placing the objective on one of the axis as suggested by the video.

Interestingly, expert users appeared to be split between two different exploration strategies. Two of the experts relied heavily on visualization tools to find patterns in the data, and then carefully selected the appropriate sampling techniques. This approach has proven to be effective for solving any trade space exploration problem regardless of the number of inputs, constraints, and objectives [40]. Other experts took advantage of the ability to rapidly generate new designs. For this single-objective problem, they used lower-dimensional visualization tools to study the progression of the objective function as they applied various sampling techniques to quickly generate thousands of designs. Regardless of the technique used, the expert users produced consistently high performance. They took advantage of the single-objective problem and generated more designs in the 10-min time frame and performed noticeably better. On the other hand, we have found no correlation between the number of designs generated and the expert's performance for multi-objective problems [40]. It is important for future work to understand how different trade space exploration strategies vary in their ability to solve problems with different numbers of inputs, constraints, and objectives.

6 Conclusions and Future Work

This research suggests that novice users with insufficient training are ineffective at using our research testbed, ATSV, in particular, and multidimensional data visualization tools, in general, to help solve a single objective, constrained design problem. In these experiments, keeping the novice users' in-the-loop did not significantly improve performance over random sampling. This can be attributed to the fact that the novices did not know how to effectively use the variety of visualization and exploration tools that were available to them, but rather used the same tools repeatedly based on their limited training and experience. Training novices to use a wider variety of visualization and exploration tools provided statistically significant improvements in their efficiency and effectiveness when compared to novices without video. In fact, the novices with video were not statistically different from the expert users, which is both a surprising and an encouraging result.

There were many limitations with this work. For this study, there was a very small population size due to participant availability, particularly the number of expert users available. A larger and more diverse sample population may have produced better and more robust results. Also, the time that the users were given to conduct the experiment was limited to 10 min. We would like to give users more time to explore the trade space to gain more insight into their knowledge discovery process. Finally, gathering users' demographic information and more formal measures of their usage and domain expertise (e.g., with optimization methods, with multi-objective optimization, with the particular problem they are solving) in future studies will enrich our analysis and strengthen our findings.

Going forward, we plan to perform Cognitive Task Analysis [12] to elicit the trade space exploration knowledge from expert users and formulate this information so that novice users can learn and benefit. This would include the information collected with the expert study presented in this paper as well as additional expert studies with different numbers of variables, objectives, and constraints. Some of this work has already been conducted comparing multi-objective problems of varying complexity [40], with other work to follow. Conducting future studies on experts can develop basic trade space exploration guidelines to be given to novice users. Shneiderman [25] states that a guidelines document can promote consistency among multiple designers. Since decision-makers are often left without an orderly approach to explore data [13], this knowledge elicitation would pertain to how to effectively use visualization and sampling tools and when these tools are most beneficial during trade space exploration. In addition to consistency, guidelines would also help promote nongoal specific problem solving, which is characteristic of expert decision-makers [19]. Additionally, it would be valuable to understand why novices pick a 'best' design when multiple competing objectives are present, as compared to how an expert would select that design. Finally, performing evaluation studies on problems with multiple objectives (constrained and unconstrained) will help us generalize our findings beyond this single objective example; however, finding suitable metrics for comparing and evaluating users' results will be critical given that multi-objective problems yield sets of nondominated (i.e., Pareto) solutions [41].

Acknowledgment

This work has been supported by a grant from the National Science Foundation (Grant No. CMMI-0620948). Any opinions, findings, and conclusions or recommendations in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

[1] Wang, G. G., and Shan, S., 2007, "Review of Metamodeling Techniques in Support of Engineering Design Optimization," ASME J. Mech. Des., **129**(4), pp. 370–380.
[2] Papalambros, P. Y., and Wilde, D. J., 2000. *Principles of Optimal Design: Modeling and Computation*, Cambridge University, New York.
[3] Balling, R., 1999, "Design by Shopping: A New Paradigm?," *Proceedings of the Third World Congress of Structural and Multidisciplinary Optimization (WCSMO-3)*, University at Buffalo, Buffalo, NY, pp. 295–297.
[4] Stump, G., Lego, S., Yukish, M., Simpson, T. W., and Donndelinger, J. A., 2009, "Visual Steering Commands for Trade Space Exploration: User-Guided Sampling With Example," ASME J. Comput. Inf. Sci. Eng., **9**(4), p. 044501.
[5] Winer, E. H., and Bloebaum, C. L., 2001, "Visual Design Steering for Optimization Solution Improvement," Struct. Optim., **22**(3), pp. 219–229.
[6] Kesavadas, T., and Sudhir, A., 2000, "Computational Steering in Simulation of Manufacturing Systems," *Proceedings of the 2000 IEEE International Conference on Robotics and Automation*, IEEE, San Francisco, CA, pp. 2654–2658.
[7] Carlsen, D., Malone, M., Kollat, J., and Simpson, T. W., 2008, "Evaluating the Performance of Visual Steering Commands for User-Guided Pareto Frontier Sampling During Trade Space Exploration," ASME Design Engineering Technical Conferences - Design Automation Conference, ASME, New York, Paper No. DETC2008/DAC-49681.
[8] Stump, G. M., Yukish, M., Simpson, T. W., and Bennett, L., 2002, "Multidimensional Visualization and Its Application to a Design by Shopping Paradigm," AIAA Paper No. 2002-5622.
[9] Stump, G., Yukish, M., Simpson, T. W., and Harris, E. N., 2003, "Design Space Visualization and Its Application to a Design by Shopping Paradigm," ASME Design Engineering Technical Conferences - Design Automation Conference, ASME, Chicago, IL, Paper No. DETC2003/DAC-48785.
[10] Simpson, T. W., Spencer, D. B., and Yukish, M. A., 2008, "Visual Steering Commands and Test Problems to Support Research in Trade Space Exploration," AIAA Paper No. 2008-6085.
[11] Seo, J., and Shneiderman, B., 2005, "A Rank-By-Feature Framework for Interactive Exploration of Multidimensional Data," Inf. Visualization, **4**, pp. 96–113.

[12] Klein, G., 1998, *Sources of Power: How People Make Decisions*, MIT, Cambridge, MA.
[13] Randel, J. M., Pugh, H. L., and Reed, S. K., 1996, "Differences in Expert and Novice Situation Awareness in Naturalistic Decision Making," Int. J. Hum.-Comput. Stud., **45**, pp. 579–597.
[14] Jiang, X., Gramopadhye, A. K., and Melloy, B. J., 2004, "Theoretical Issues in the Design of Visual Inspection Systems," Theor. Issues Ergon. Sci., **5**(3), pp. 232–247.
[15] Petre, M., and Green, T. R. G., 1993, "Learning to Read Graphics: Some Evidence That 'Seeing' an Information Display is an Acquired Skill," J. Visual Lang. Comput., **4**, pp. 55–70.
[16] Zhu, Y., 2007, "Measuring Effective Data Visualization," *Advance in Visual Computing*, Springer, Berlin, pp. 652–661.
[17] Simpson, T. W., Spencer, D. B., and Yukish, M. A., 2008, "Visual Steering Commands and Test Problems to Support Research in Trade Space Exploration," *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Victoria, British Columbia, Canada, AIAA, September 10–12, Paper No. AIAA-2008-6085.
[18] Orasanu, J., and Connoly, T., 1993, "The Reinvention of Decision Making," *Decision Making in Action: Models and Methods*, G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok, eds., Ablex Publishing, Norwood, CT, pp. 3–20.
[19] Gilmour, P., and Corner, J., 1998, "The Role of the Expert's Decision Making Skills in Management," 33rd Annual Operational Research Society of New Zealand, Ablex Publishing, Auckland, pp. 195–204.
[20] Bouwman, M. J., 1984, "Expert vs Novice Decision Making in Accounting: A Summary," Account., Organ., Soc., **9**(3/4), pp. 325–327.
[21] Anderson, J. R., 1982, "Acquisition of Cognitive Skill," Psychol. Rev., **89**, pp. 369–406.
[22] Ritter, F. E., and Bibby, P. A., 2008, "Modeling How, When, and What is Learned in a Simple Fault-Finding Task," Cogn. Sci., **32**, pp. 862–892.
[23] Wiedenbeck, S., 1986, "Organization of Programming Knowledge of Novices and Experts," J. Am. Soc. Inf. Sci., **37**(5), pp. 284–299.
[24] Jones, D. G., and Endsley, M. R., 1996, "Sources of Situation Awareness Errors in Aviation," Aviat., Space Environ. Med., **67**(6), pp. 507–512.
[25] Shneiderman, B., 1998, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Addison-Wesley, Reading.
[26] Buja, A., McDonald, J. A., Michalak, J., and Stuetzle, W., 1991, "Interactive Data Visualization Using Focusing and Linking," *Proceedings of the IEEE Conference on Visualization*, IEEE Society, San Diego, CA, pp. 156–163.
[27] Becker, R. A., and Cleveland, W. S., 1987, "Brushing Scatterplots," Technometrics, **29**(1), pp. 127–142.
[28] Stump, G. M., Yukish, M. A., Martin, J. D., and Simpson, T. W., 2004, "The ARL Trade Space Visualizer: An Engineering Decision-Making Tool," 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA, Albany, NY.
[29] Law, A. M., and Kelton, W. D., 2000, *Simulation Modeling and Analysis*, McGraw-Hill, Boston, MA.
[30] Price, K., Storn, R., and Lampinen, J., 2005, *Differential Evolution - A Practical Approach to Global Optimization*, Springer-Verlag, Berlin.
[31] Robic, T., and Filipic, B., 2005, "DEMO: Evolution for Multiobjective Optimization," *Third International Conference on Evolutionary Multi-Criterion Optimization*, Springer, Guanajuato, Mexico, pp. 520–533.
[32] Stump, G. M., Simpson, T. W., Yukish, M., and O'Hara, J. J., 2004, "Trade Space Exploration of Satellite Datasets Using a Design by Shopping Paradigm," 2004 IEEE Aerospace Conference Proceedings, IEEE, Big Sky, MT.
[33] Neilson, J., and Mack, R. L., 1994, *Usability Inspection Methods*, Wiley, New York.
[34] Wagner, T. C., and Papalambros, P. Y., 1991, "Optimal Engine Design Using nonlinear Programming and the Engine System Assessment Model," Ford Motor Company, Technical Report No. SR-91-154.
[35] McAllister, C. D., and Simpson, T. W., 2001, "Multidisciplinary Robust Design Optimization of an Internal Combustion Engine," ASME J. Mech. Des., **125**(1), pp. 124–130.
[36] ANSI/ASME PTC 19.1-1998, *Measurement Uncertainty, Part I*, ASME, New York, NY.
[37] Wolf, D., Simpson, T. W., and Zhang, X., 2009, "A Preliminary Study of Novice and Expert Users' Decision-Making Procedures During Visual Trade Space Exploration," ASME Design Engineering Technical Conferences—Design Automation Conference, ASME, San Diego, CA, Paper No. DETC2009/DAC-87294.
[38] Devore, J. L., 1995, *Probability and Statistics for Engineering and the Sciences*, Wadsworth, Belmont, CA.
[39] Booth, T., 1967, *Sequential Machines and Automata Theory*, Wiley, New York.
[40] Wolf, D. R., 2009, "An Assessment of Novice Users' Decision-Making Strategies During Visual Trade Sapce Exploration," M.Sc. Thesis, Mechanical Engineering, Penn State University, University Park.
[41] Wu, J., and Azarm, S., 2001, "Metrics for Quality Assessment of a Multiobjective Design Optimization Solution Set," ASME J. Mech. Des., **123**(1), pp. 18–25.