Proceedings of the ASME 2021
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC/CIE2020
August 17-20, 2021, Virtual, Online

DETC2021-71830

CONNECTING DESIGN ACTIONS, REASONING, AND OUTCOMES IN CONCEPT-SELECTION

Yakira Mirabito
Dept. of Mechanical Engineering
University of California, Berkeley
Berkeley, CA, USA
Email: yakira.mirabito@berkeley.edu

Kosa Goucher-Lambert
Dept. Mechanical Engineering
University of California, Berkeley
Berkeley, CA, USA
Email: kosa@berkeley.edu

ABSTRACT
Final concepts are often not the most creative or innovative design within the solution space. The purpose of this research is to gain insight into the decisions made in concept selection. In particular, we studied how designers link multiple decision-making elements together, including: actions (what people do), reasoning (why they do it), and design outcomes (an objective measure of engineering performance). Fifty-seven participants were tasked with solving a design challenge relating to a robotic gripper by selecting a design within a predefined design space. Each design had a corresponding measure (termed "success rate") which enabled each designer's performance to be quantified and compared against other designers. The task was hosted on an interactive interface in which design actions were collected. A post-task survey probed for the reasoning behind design actions. Characterization of decision-making behavior and reasoning was rooted in prior design literature. Design actions were quantified concerning the degree of design space explored and the decision-making strategies employed. Key results include design strategies such as manipulation techniques, the impact of maximum observed success rates, and a willingness to submit an alternative solution which influenced design outcomes. Although designer preferences validated the design strategies identified, there was no correlation between the decision factors considered and improved outcomes. The methods and findings from this work assessed the underlying dynamics when engineers selected less innovative or creative solutions and recommended decision-making strategies that should be considered to improve design outcomes.

1. INTRODUCTION
Engineering design is an iterative decision-making process that engineers use in developing products and systems. An ample design space of possible solutions is desirable [1,2]. There are significant efforts in the design research community to increase engineers' innovative capabilities in the concept generation phase [3]; however, engineers often select less innovative solutions [4]. This gap continues to grow as the range of possible solutions increases as designers' skills improve and as generative designs become commonplace [5]. Designers select concepts with lower engineering outcomes, although more optimal concepts are within reach, creating a tension worth further investigation [6–8]. Thus, there is a need to better understand the decision-making process of concept selection. This paper investigates the design behavior and reasoning designers used in selecting a final design concept to determine what factors enabled high-performing designs.

Research in decision-making in engineering design and, more specifically, concept selection is critical since these designs are the designs that advance into the later stages of development. These final designs may not be the "best," most innovative, or creative, but these are the concepts pursued. By understanding what factors enable high-performing design, the design research community can help engineers utilize specific strategies that improve design outcomes, such as systematic exploration. Prior work has cited feasibility as a barrier in selecting more creative or innovative solutions [6,7]. Feasibility is a concern across students and industry professionals since resources are finite [4,9]. For example, instilling specific considerations or addressing concerns that hold designers back from selecting highly creative or innovative designs can better target designers who exhibit similar behavior.

A two-part human subjects study, a design task and a post-task survey, has gathered information on design actions and reasoning. Design actions have been defined as the observable and quantifiable steps a designer takes within the task. Design reasoning has been defined as the rationale, motivation, and preferences participants use to explain their design actions. Relationships between actions, reasoning, and engineering performance have been determined and analyzed. This paper presents these findings and closes with future directions on predicting design actions based on reasoning provided or vice versa. The ability to derive such conclusions with only design action or design rationale data will provide further insight into past research.

2. BACKGROUND

2.1 Concept Selection

The engineering design process involves iterating between divergent and convergent steps. Evaluating a concept is a crucial stage that converges on fewer concepts than initially generated [10]. This decision is an executive function that involves reasoning and action [11]. The design research community's efforts aim to increase early-stage design to allow for more creative or innovative solutions [1,2]; however, research shows engineers often fail to select such designs despite higher-performing designs within their solution space [4].

Ulrich mentions that all teams use some method to choose among concepts regardless if the method is stated [10]. Concept-selection tools such as decision matrices vary in their effectiveness [12,13]. A Pugh decision matrix lists the concepts to be selected between, then the design team (individually or as a whole) rates the concepts on a series of qualities the team deems most important [14,15]. Although the tool focuses on objectivity, literature has identified cases where team members have selected criteria to rate that would help confirm and build support for concepts of their preferences [16]. Due to the various components of decision-making, the paper herein focuses on two aspects, design actions which are the observable and quantifiable steps in selecting a concept, and reasoning, which are the explanations or motivations behind why a design action. Design reasoning is not necessarily observable unless an engineer directly provides the rationale behind an action.

2.2 Design Actions

Design actions are the incremental and observable steps that occur in the design process. Design exploration/search, testing, and final concept submission are the design actions of interest in this study. They are measurable interactions with the study interface which identifies the sequential path taken by each designer. The engineering design community has stated that a solution is the latest version of an idea that has undergone many transformations [17]. The idea begins as the initial idea state that then changes into the final goal state using transformations regarding its functions, features, or materials. Hay and colleagues [18] have noted that the limitations of idea transformations have constraints, bounding the solutions' space. These idea transformations can be characterized in numerous methods.

Goel has introduced a two-dimensional representation of idea transformations that are lateral or vertical [17]. Lateral means changes in the design, while vertical represents the level of detail in a design. This interpretation provides a sense of structure to the rather vague term of idea transformations. Transformations allow more insight in comparison to final outcomes in the design process. Instead, by characterizing each step and viewing all the steps holistically, insights surrounding the path (via search/exploration), strategies, and the degree to which they agree with prior findings can be quantified and connected to existing frameworks.

2.2.1 Visualizing and quantifying design actions

Design actions in concept-generation have been visualized through Linkography [19], Interaction Dynamics Notation [20], time series cartesian graphs, state transition diagrams, and force dynamics notation. Linkography is a protocol analysis method that requires the researcher to code the steps and then determine the connections between steps. Methods that require qualitative coding of design actions introduce researcher subjectivity which then requires systematic inter-rater reliability or researchers recoding until they agree on codes. Such methods attempt to go beyond the objective steps a design takes and introduce team dynamics or the connections between steps. The methods for visualizing concept-generation processes have inspired the visualizations for concept-selection processes in this study.

2.2.2 Frameworks for assessing design actions

Existing frameworks in information retrieval, optimization, or rational decision-making, are approaches to analyze design actions in the concept-selection process. Moving beyond final design task outcomes to the moment-to-moment characterization of design actions aims to distinguish and explain strategies or sequential patterns that appear across multiple designers. A combination of the following frameworks inspired the analysis techniques implemented in this work.

The *theory of information foraging* from human-computer interaction research explains how users search for information using a digital interface. This theory is mentioned since the design task for this research study used a digital interface. The design actions on the digital interface have been modeled after design actions a designer would take in the physical world. According to information foraging theory, information can be sought, gathered, shared, and consumed analogously to animal foraging evolutionary models [21]. In concept selection, information is necessary to decide which design to advance.

Information foraging introduces a framework to characterize design actions concerning depth and breadth. Engineers seek out information on concepts they are deciding on via background research, interviews, prototyping, or testing. Searches can be considered as the exploration of sources of information. Information foragers then judge or assess the potential source. Foraging behavior has been further classified into breadth and depth search. Breadth is defined as high-level information seeking or consuming information for short periods. In comparison, depth involves spending a long time processing information or in-depth information details. Exploration breadth is the number of sources of information, while exploration depth is the number of searches within a source [22]. Design space explored or the design actions performed relate to this depth and breadth.

Optimization techniques use systematic methods to solve a problem [23]. Sequential path characterization can be thought of in deterministic or stochastic frames. In a deterministic mindset, an engineer uses *single or multi-attribute manipulation* strategies. Single-attribute manipulation means altering one design parameter at a time to improve desired outcomes. In comparison, multi-attribute manipulation translates to altering multiple parameters at a time to arrive at the desired solution. Stochastic approaches can be used to describe the randomness or unexpectedness behind some decisions. In this study, testing strategies from participants have been described by parameter tuning behavior.

Decision-making theories often fall into either a rational or psychological model [24]. Decisions involve action and cognition. A decision is a reaction to a situation involving judgment, expectations, and evaluation [24]. Engineering has traditionally followed more rational approaches on how an engineer makes decisions. Meanwhile, the psychological approach highlights how people actually behave. Ulrich's design process [10] aligns well with the *rational decision-making model*; both are cyclic and overlap in problem identification, criteria, concept generation, evaluation, and implementation. However, rational models fail to account for the irrationality of human nature [25]. *Bounded rationality theory* introduces limitations to the rational decision model resulting in engineers exhibiting satisficing behavior—where the first design that meets the criteria is selected rather than maximizing engineering outcomes by exploring more of the design space [23]. These limitations can be attributed to human and environmental factors imposed on the decision-making process. Those factors are the aspects an engineer considers or reasons within their decision-making process.

2.3 Design Reasoning

Design reasoning, in this paper, is the why behind a design action or the factors influencing decision-making. These reasons vary from concrete, such as prototype feedback, to more abstract means such as personal preferences or values. Reasoning, as discussed herein, focuses on the conscious considerations influencing the overall design task. It is essentially what a designer thinks about or leverages as the rationale behind their design actions. The factors that have been analyzed stem from design reasoning literature and protocol studies, such as testing or previous experience [15,18,26]. Prior work has found that designers with more years of experience rely more heavily on this experience or knowledge than those with less experience [27,28]. Design reasoning can be stated explicitly, but implicit design reasoning is likely occurring.

2.3.1 Collecting design reasoning

Design reasoning involves the explicit and implicit understanding of cognition. These cognitive processes can be gathered through think-aloud studies, interviews, or surveys, which are standard methods in which researchers directly ask designers to share their thought processes or rationale behind their design actions. Individuals may censor or modify their reasoning. Additionally, implicit rationale such as cognitive biases behind design actions may not be verbalized or recognized by the engineer. The research team is then responsible for coding or interpreting implicit reasoning.

2.3.2 Frameworks for situating reasoning in the design process

Reasoning frameworks, such as information seeking [21] or rational theories, explain why users perform specific actions. The rational model links logical reasoning to design actions objectively, considering what engineers do. Although engineers are often considered rational actors, they routinely behave irrationally, which indicates their reasoning is likely irrational. Constraints and uncertainty lie relatively in between rationale and psychological models. With new limitations, the decisions appear irrational. Work from Hazelrigg describes all engineering design as suboptimal due to the different objective measures individual designers have used for their process [25]. That means what is rational for one engineer is irrational for another. Thus, each engineer on a team optimizes different metrics (external and internal), and no design truly achieves optimality in all engineers' eyes. Furthermore, psychological factors have been proposed as explanations for decisions that deviate from rational methods, including social dynamics, culture, and beliefs.

Design research speculates on the influence of cognitive biases in engineering design [4], and a few papers have identified the presence of biases such as anchoring [29], bias against creativity [30], and confirmation bias [16,31]. The literature mentioned detects these biases through qualitative studies of small samples, but the quantified influence of cognitive bias on engineering outcomes has yet to be considered. Moreover, visual interface designs are linked to specific search patterns and behaviors [32]. Research from Karim and colleagues link clockwise and anticlockwise behavior with demographics. This patterned behavior is called a directionality bias in visuospatial functioning. Together, rational and psychological factors are used to explain design reasoning.

2.4 Approach

This paper aims to combine design actions (e.g., viewing designs, testing, selecting concepts) and design reasoning (e.g., engineering principles, previous experience) to understand their influence on one another and a participant's performance within the concept-selection stage of engineering design. The overarching question of this research is: how are design actions, reasoning, and outcomes linked? A design space with predefined performance measures has been utilized to answer the research question. Participants have been grouped by design outcomes from their final concept selection. The similarities and differences of design actions and reasoning of participant groupings have been identified.

3. MATERIALS AND METHODS

3.1 Participants

A total of 68 participants were recruited for the design study using a call for participation at the University of California, Berkeley. Participants were compensated $10 for their participation for a total of 30 min. A bonus of up to an additional $20 was offered contingent on task outcomes. Participants were screened and required to be 18 years or older with engineering or design experience to participate in the research study. Experience ranged from completing one design class upwards to 10+ years as an engineer in industry. All participants read, agreed, and signed a consent form. Participant demographics included undergraduate, graduate, and working professionals with backgrounds in engineering and the sciences. Data from 57 participants (30 men, 26 women, one non-binary person) was used for data analysis. Data from 11 participants were removed due to a lack of following instructions and equipment errors.

3.2 Experimental Design

The experiment took approximately 30-minutes and consisted of two parts: the design task and a post-task survey. Participants were tasked with designing a gripper surface for a robotic arm, as shown in Figure 1. The context and task information shown in the study was copied below.

Context: You are an engineer on a robotics team tasked with designing a gripper surface for a robotic arm, as shown below. A dishwashing robot uses the grasper in a wet and slippery environment due to dish soap. The design's success is determined by the robot's success rate in grasping a range of dishware.

Task: Your team created 21 possible solutions based on 7 surface designs and 3 material hardness options. They have created a data sheet for each design, and now it is up to you to decide which one to move forward with for production. You have 10 minutes to complete the task and have the option to test up to 5 designs to see their success rate.

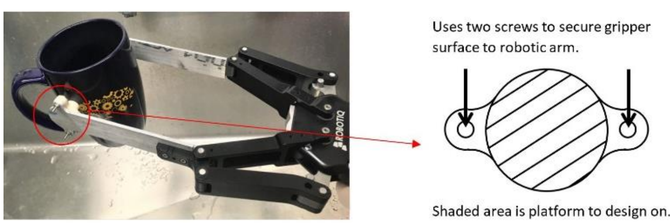

FIGURE 1: Image from robotic gripper study [33] and diagram created to relate design options to the testing procedure

Participants received the design task instructions and information on how the gripper design was tested: the success rate in picking up objects of various shapes and textures (the exact dimensions were not provided). Each participant searched for the optimal design within a predefined design space of gripper surfaces provided by colleagues at UC Berkeley [33]. The designs shown to each participant were the same, and two parameters could be manipulated (2D design, material hardness), although additional parameters were provided on the datasheet. The purpose of the data sheet was for participants to identify design features that changed. Contact area remained constant, while feature number, radius, and hardness varied, impacting bending stiffness and a design's performance. The task had a 10-minute time limit and five tests maximum to see a design's success rate. A five test limit was set to mimic real-life constraints in the design process where a limited number of designs can be tested due to time or financial constraints. Preliminary experiments found that few participants converged on a "good" design with less than five tests. Based on the design's success rate of the final design submitted, participants earned up to an additional \$20—twenty dollars for the best design, between \$5 and \$15 designs over 50%. Participants interacted with an interface for the design task portion, hosted on Figma, which displayed the consent form, task instructions, and all possible design options, as shown in Figure 2.

The time and number of tests conducted were left to the participant to monitor due to interface constraints. However, this decision to self-monitor was aligned with what engineers and designers experience outside of controlled studies where they are expected to meet deadlines and stay within budget. Although the robotic gripper design could be optimized using a computer program, this predefined solution space removed researcher subjectivity in classifying a participant's design actions and performance that alternative experimental setups may have introduced (e.g., having participants sketch their designs followed by researchers rating designs using rubrics). Design actions increased or decreased a participant's overall performance. The distance between any two designs' success rates was computed and used for analyses.

The second part of the study had multiple-choice and open-ended questions via Qualtrics about their decision process used and the factors considered. Open-ended questions surrounding a participant's overall, initial, and final approaches were captured to provided reasoning in the participant's own words. Afterward, participants were asked to rate eight factors, based on prior literature, that influenced their concept-selection process. Additional questions expanded on the factors selected, such as the datasheet, engineering principles, and previous experience. Any factors they rated had some influence on their decision were then carried forward and provided in a list for participants to rank. To assess the impact that authority bias had on design decisions, the study provided participants with an option to submit an alternative design from a fictional senior robotics engineer. That design had a success rate of over 50% but was not the most successful design. Participants then rated their confidence in their design and the senior engineer's design on a five-point Likert scale.

3.3 Analysis

Data collected from the Figma website using Maze.co, recorded each screen a participant visited and exported the duration, screenId, and sequential path for each participant. The exported task data from Maze and survey data from Qualtrics removed identifiers, were cleaned, and merged for analyses. Design actions defined in this study were steps traveled within the solution space and used objective methods to evaluate a design's performance [18]. Design reasons defined in this study are the factors considered in the concept-selection task.

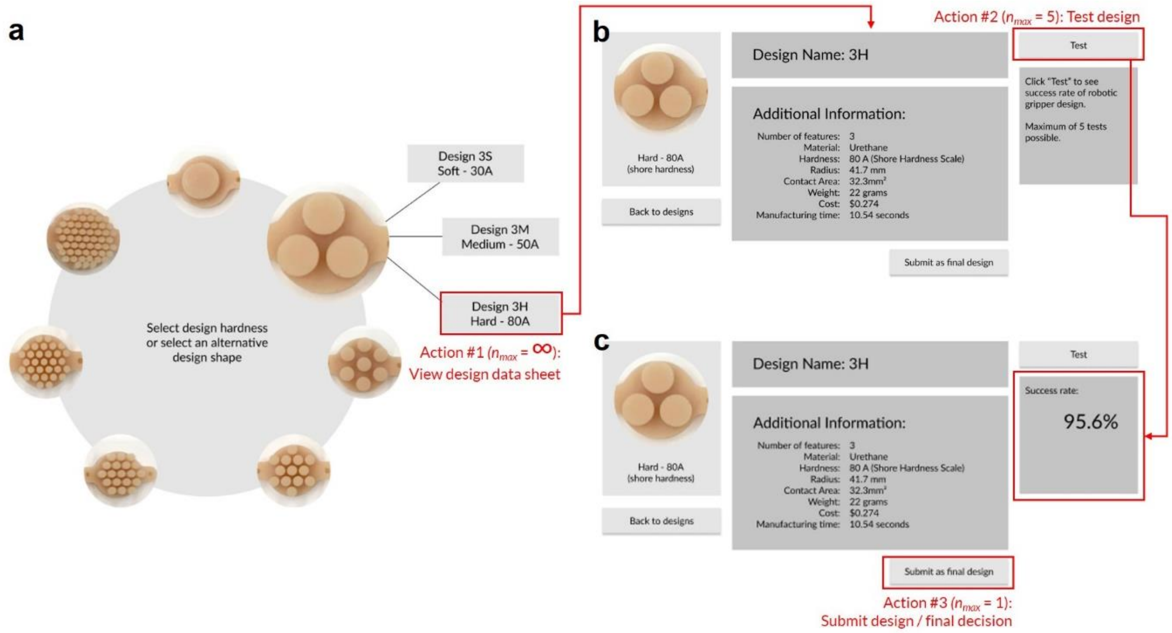

FIGURE 2: The design actions of interest are represented as actions 1-3, and n_{max} indicates the number of times said action could occur. (a) Each of the seven geometries (shown as images) branched out to include a soft, medium, and hard version. Clicking one of the branched-out designs opened a datasheet (b) for that design. A test displayed a design's success rate (c).

3.3.1 Design Actions

For this study, the goal was to submit a design of the participant's choosing to move forward to production, the next step in the fictional robotics team's design process. Participants explored/searched among the seven design geometries and three material hardness combinations (Figure 2a). After clicking on a design, a screen showed a data sheet for the design where participants could test a design and see the success rate (Figure 2b). Concerning information foraging theory, the range of sources included all options screen, navigation of design geometries with material options branched-out, datasheet information, and lastly, the design's success rate (Figure 2c).

Design actions were classified by the sequential screens participants visited. The actions quantified in this study are listed in Table 1, and for the analyses of design actions, only actions C, D, and E, were considered since they each have 21 possible options. The designs explored (total and unique) were determined along with the time spent on each. Moreover, testing behavior was gathered for each participant regarding the tests conducted and their order. Whether a participant carried out single or multi-parameter manipulation decisions was concluded per participant and on an action-to-action level using screens-visited.

3.3.2 Factors influencing decisions and performance

Design reasons were gathered and analyzed from the survey component, which used open-end and multiple-choice style questions. The importance ratings and rankings relative to one another were computed, and statistical tests were used to determine the findings' significance. Pearson's correlations were run to assess relationships between factors considered and their influence on a participant's performance. Confidence and authority bias were two elements that were explored in the design task after a participant selected their final design. Participants' willingness to submit an alternative design was computed using their final design's success rate and self-reported confidence level.

4. RESULTS

The following results explore the relationships between design actions, reasoning, and outcomes. The study asked participants to select and submit a design from 21 possible options. Participants searched and explored the design space within the 10-minute limit and tested up to five designs to see the design's success rate [33]. Note that lowercase (n) refers to design actions and an uppercase (N) to participants. The 57 participants made 2451 design actions; 12% of actions were test actions (n=282).

TABLE 1: Design Actions for all participants combined (n = 2249). Excludes actions related to reading instructions/consent.

Design Actions	*# of actions (n)*
A. View all options	*83*
B. Click design geometries	*1218*
C. Visit datasheet for a design	*609*
D. Test a design	*282*
E. Submit as final design	*57*

Participant groupings were determined using design outcomes measured by the design's success rate, as indicated in Figure 2c. Participants in the *high-performing* category (N=19) achieved the best possible design (95% success rate). The *average-performing* category achieved success rates above 50%, excluding the optimal design (N=20), while those in the *low-performing* category achieved success rates under 50% (N=18). The following sections use the participant groupings to identify differences and similarities between the groups' design actions and reasoning. Figure 3 and Figure 4 remove the *average-performing* group to focus on the extremes of participant performance.

4.1 Design actions differed by groups, and a select few strategies resulted in improved outcomes.

Datasheets visited, tests conducted, and designs submitted were the primary design actions analyzed. The heat map in Figure 3 shows the number of design actions per design for the 19 high-performers (top) and 18 low-performers (bottom). The darker shading indicates designs that were explored more than the lighter shaded regions. Thus, for the top heatmap, "Design 3 Hard" was explored a max of 56 times, while for the bottom heatmap, "Design 55 Soft" was explored a max of 29 times. For context, the horizontal axis shows the seven design geometries, while the vertical axis shows the material harness options (soft, medium, hard). The percentage overlaid indicates the success rate for that design. The different heatmaps for each group show apparent differences in their design behavior.

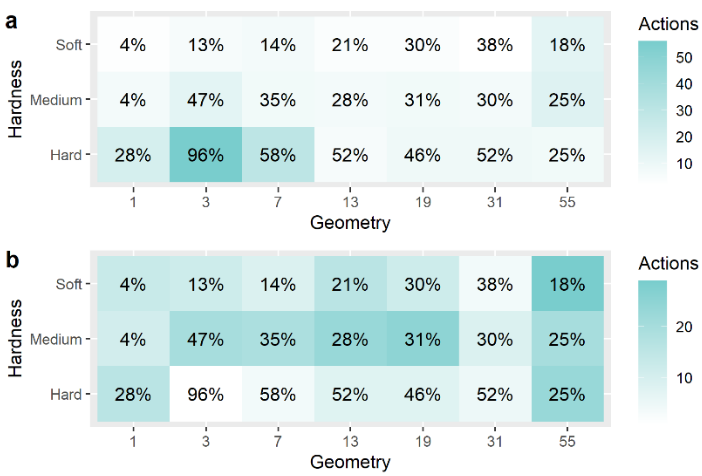

FIGURE 3: Two-dimensional representation of design space explored. Shading represents the number of design actions for that design. (a) High-performing designers and (b) low-performing designers. The success rate for each geometry-hardness combination is overlaid as a percentage.

4.1.1 Increased design space exploration did not increase design outcomes.

The scatterplot in Figure 4 shows one quantification of the breadth and depth of the 57 participants' design actions. Breadth was the number of datasheets they visited (percent of design solutions explored), while depth was the success rates of design submitted. Multiple models attempted to explain the data shown; however, they had low R^2 values with statistically significant p-values. Although the models could not explain the variation of data, a slightly negative correlation between the design space explored and the final design's success rate was observed *(r = -.27)*. Therefore, exploring more of the design space did not improve final success rates and appeared to have a negative influence.

Alternative analyses of breadth included the total number of screen visits and the total task time. Correlational statistical test between 1) the total number of screen visits (including multiple visits to datasheets) and success rate of final design and 2) total task time and success rate of the final design are not significant. The relationships between what participants did were inconclusive for this study using this version of breadth and design outcomes.

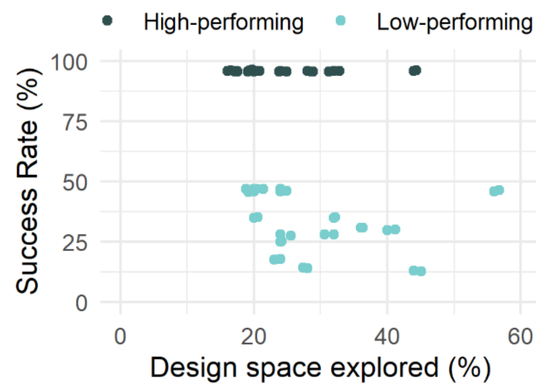

FIGURE 4: Scatterplot of the percentage of design space explored (number of datasheets explored of the 21 possible datasheets) and the success rate of the final design submitted. Each participant is represented as a single data point.

4.1.2 Single parameter manipulation as a testing strategy improved engineering outcomes.

Design geometries and material hardness values were the two directly adjustable parameters. Keeping the geometry constant while changing hardness or vice versa was a single parameter change. Changing both design geometry and hardness were multi-parameter changes. Table 2 shows testing sequences from two participants, one with a single parameter technique and one with a multi-parameter technique. The 19 high-performing participants used single-parameter manipulation most among the participant groupings.

Participants who changed one parameter at a time between tests were expected to have higher engineering outcomes than those who changed two parameters at once. Eight out of 19 *high-performers*, 7 out of 20 *average-performers*, and 1 out of 18 *low-performers* strictly used single-parameter manipulation. Although 11 of the 19 high performers used multi-parameter techniques, they primarily used a single parameter technique, and only after they achieved the optimal design did they test a completely different design. A chi-squared test was run to determine whether the techniques (single or multi) employed by each grouping were significant or due to chance. The results show a statistical significance between the three groups ($\chi^2 = 6.848$, $p < .05$), meaning that high performers' testing strategies are statistically different from average and low performers. A one-tailed Mann-Whitney test assessed whether those who carried out single parameter manipulation had improved success rates (average of the group) compared to those who used multi-parameter manipulation. The results are statistically significant (Mann–Whitney $U = 198$, $n_1 = 16$, $n_2 = 41$, $P < .05$ one-tailed). There was a 17% increase in performance for those who used single parameter manipulation. The open-ended questions from the survey confirmed that participants used these strategies.

4.1.3 Two-thirds of participants selected the design with the maximum observed success rate.

Submitting a final design with the max observed success rate from testing was the expected design behavior. Results show that 72% of participants *(N=41)* behaved as expected while 28% of participants *(N=16)* behaved unexpectedly and did not submit the design with the best-observed success rate. Of those 16 participants, ten were below the 50% threshold needed to earn a bonus, while six were already above the 50% threshold. Upon closer inspection, of those 16 participants, 44% *(N=7)* took a risk on an alternative design not tested and achieved a higher design outcome, while 56% *(N=9)* took that risk and resulted in lower design outcomes. Of the six participants above the 50% threshold and risked submitting an untested design, only one person outperformed their previous design while the other five participants not only resulted in lower performance designs but no longer qualified for any bonus. Those who submitted designs with the max observed success rate had a 25% mean difference in success rates (69% for max observed and 44% for untested designs).

TABLE 2: Manipulation techniques with examples sequences and average participant performance

Technique (example sequence)	*Participants*	*Mean*
Single parameter (13M →55M →7M →7H →3H)	*N=16*	***74%***
Multi-parameter (55S →1H →19M →3S →55H)	*N=41*	***57%***

4.1.4 Willingness to submit an alternative solution was influenced by confidence levels and authority bias.

Another aspect of the study offered participants the option to change their final design submission to the design submission of a fictional senior robotics engineer. Those who submitted their own design *(N=41)* had, on average, a success rate of 69% and a confidence rating of 3.31. The confidence rating was on a 5-point Likert scale. Those who submitted the senior engineer's design *(N=16)* had, on average, a success rate of 45% and a confidence rating of 2.19. One-tailed Mann-Whitney tests between the two groups for mean success rate (Mann–Whitney $U = 150$, $n_1 = 41$, $n_2 = 16$, $P < .05$ one-tailed) and mean confidence (Mann–Whitney $U = 132$, $n_1 = 16$, $n_2 = 41$, $P < .05$ one-tailed) were statistically significant. Thus, participants who accepted the suggested solution had lower confidence and lower success rates from testing.

4.2 Tests, guesses, and engineering principles were the three most important decision-making factors.

Design justifications were expected to help explain the design actions discussed in Section 4.1. This section presents factors and their influence on the design task and outcomes. Factors were collected via the survey in both close-ended and open-end questions. Table 3 shows the average ratings for the eight primary factors of interest. Open-ended responses validated design action strategies observed in the previous section and are discussed in those respective discussion sections.

The factor analysis evaluated the importance of all participants' factors. A one-way ANOVA was used to assess statistical differences among the importance ratings for each factor considered *($F(7,448) = 74.1$, $p < .001$)*. A follow-up TukeyHSD was conducted where most factor pairs were also significant, which indicated that each factor played different levels of importance. Pearson's correlational tests were conducted between participant's ratings of factors (i.e., Tests) and their final success rate (total of eight tests); The correlation coefficients were low or negligible in the range of $r = -.10$ to $.23$, which means prioritizing certain factors did not improve success rates. Pearson's correlation was also run for each pair of factors to understand how factors correlated with other factors. Only previous experience and engineering/design principles had a moderate correlation *($r = .47$)*, while all other pairs had lower correlation values (r). For example, guesses and previous experience were expected to be negatively correlated since more knowledge should have signaled less need for guesses, but no such correlation existed.

TABLE 3: Factors rated by importance in concept-selection

Factor	*Mean rating*	*Standard error*
Tests	*4.81*	*.073*
Guess	*3.75*	*.131*
Principles	*3.25*	*.123*
Previous Exp	*2.82*	*.174*
Datasheet	*2.70*	*.168*
Time	*2.32*	*.168*
Financial	*1.42*	*.086*
Aesthetics	*1.42*	*.100*

5. DISCUSSION

A human subjects experiment used a predefined solution space with designs of varying success rates to gain a deeper understanding of design actions and reasoning in concept selection. The participants' groupings by engineering performance paved the foundation for which similarities and differences between the groups' design actions and reasoning were analyzed. Excluding the average-performance grouping from Figure 3 and Figure 4 directs attention to the extremes to show the contrast between observed behaviors of low and high performers. Through this multi-attribute understanding of decision-making, the hope was that the design research community could better understand the dynamics influencing concept-selection decisions and, ultimately, design outcomes.

5.1. Multiple design actions led to high-performing designs.

The degree of design space search/explored, manipulation techniques used, expected behavior observed, and willingness to change designs influenced engineering outcomes in this study. Design actions of interest were viewing design datasheets, testing, and submitting a design as visualized in Table 1 and Figure 2, ranging from high-level information to in-depth information.

Exploring more of the design space did not lead to improved success rates. Therefore, participants who viewed more of the designs did not do better than those who explored a smaller portion of the design space. This observed finding contradicts previous literature, which has found that exploration improved design innovativeness [34]. This contradiction may be due to the performance measures used, innovation metric versus success rate. Information foraging theory complements the findings and states that users who searched less (% of design space in this study) but at more profound levels attained higher information values [21,22]. In contrast, those who explored more for brief times acquired less valuable information or lacked understanding of the design challenge. Higher values of information gain in this research assumed that participants spent more time on fewer designs because they knew that information was more valuable than alternative designs or were more meticulous in their design actions than other participants.

Alternatively, the degree of the design space explored showed how exploratory a person was, the degree of design fixation experienced [29], or the degree of uncertainty a participant felt [25]. Those who searched more might be uncertain about the design outcomes due to their lack of systematic testing or limited engineering knowledge/experience. In reviewing the qualitative survey data, one participant "*knew that harder grippers would have higher success rate[s]*," while others may have arrived at that spot by luck or instinct as another participant said, "*it was design intuition to refer to a hard material with a triangular point force-based design.*"

Moreover, when evaluating tests conducted ($n_{max}=5$), single parameter manipulation resulted in improved outcomes. This strategy was more successful than a trial-and-error approach that rarely resulted in high success rates. These results indicate how these designers and engineers might select concepts in other situations. The findings do not reveal why participants used manipulation versus trial-and-error methods. That reason is likely related to designer experience but may occur because they jumped into the task too quickly and used up their tests early on. A few participants realized their error, but it was too late, such as this participant, "*I mistakenly chose to try to test for the best hardness and surface area at the same time, which did not work out favorably. I should have focused on a single aspect.*" Differences in when and how certain groups tested align with Tahera and colleagues' findings, indicating that testing actions and reasoning of testing vary considerably [35]. Future inquiry in understanding the nuances of testing behavior in this study will be necessary.

Most participants (72%) behave as expected and selected the design with the maximum observed success rate. This conclusion suggests that designers who did not submit the design with the max observed rate placed the interest of the fictional design team lower than the financial incentive offered (designs over 50%). Final decisions that did not align with expected decision-making behavior highlight a design engineer's willingness (in this study conditions) to send a design to production with more unknowns in the hopes of better performance than a known design of average performance. Of course, a financial incentive for the decision-maker could be a positive aspect for one individual. The results suggest that one's self-interests were prioritized. This decision behavior might occur in industry or the classroom by these same designers. Meanwhile, the decision to submit the design that met the threshold may point to satisficing behavior [36]. Although the aspect of a bonus was carefully introduced and piloted before testing, possible influences may have been overlooked [37]. Examining scenarios outside of the lab where bonuses are involved may be of future interest in engineering to determine if lower-performing designs with many unknowns would result.

Confidence levels and authority bias influenced a designer's willingness to submit a fictional senior robotics engineer's design. Considering 16 participants (with mean success rates of 45%) changed their design submission, this implies similar behavior may occur on engineering teams in industry. The willingness to use the other person's design depended on a participant's max observed success rate and self-confidence. The literature on cognitive bias has identified authority bias as a factor that influenced decision-making [38,39]. Design research from an architecture course detected the high degree of influence instructor feedback had on project direction, even when the current direction was quite successful [40]. The implications for real-world scenarios can be detrimental to the team when low-performing recommendations are pursued over high-performing designs from a junior designer or designer with low confidence in their design. These results support the need for structured concept-selection methods or tools to reduce these influences, else low-performing designs could result.

5.2. Reasons for concept-selection approaches were rooted in various factors.

Both expected and unexpected design reasoning were identified in the post-task survey. Testing, guesses, and principles of engineering/design were the top three factors ranked. However, the factors considered were not correlated with design outcomes. Low correlations between the factors and success rates point to a disconnect between individual values and translating that value to a design approach. For example, both high and low-performing designers rated testing with high importance. However, after examining their design actions, high-performers used test feedback strategically with single-parameter behavior while low-performers relied on trial-and-error.

Additional factors not asked as close-ended questions were identified from the open-ended responses. Patterns of isolating and prioritizing variables from the data sheets were identified as the reasoning behind the manipulation techniques discussed. Multiple participants referred to three contact points or triangular shapes as a rationale for starting or testing the design with three features that stemmed from engineering principles or prior experience. A few participants appeared to have used analogical reasoning between the task at hand and their prior experience/ engineering knowledge to support their decision. These results suggest that although participants in this task documented reasons for their actions, the same rationale did not lead to the same approach. Design research literature has determined that analogical reasoning and analogical distance were correlated with improved design outcomes for idea generation [41]. Perhaps future work should consider the impact of analogical reason on concept selection. Despite motivations and reasons for completing a task were aligned, the approaches used and designs tested differed.

The surprising element in the open-end responses was the words participants used to communicate their rationale. Participants' responses point to the known challenge that students often struggle to articulate their decisions during the design process and their justifications for doing so [42,43]. Statements surrounding justification may be rhetorical strategies to strengthen their own beliefs. For example, one person used intuition that a more rigid surface was better than the other hardness options. However, the basis for their belief was not stated. Another person assumed that hard would be better and another hypothesized that hard would be better. Using intuition, hypothesizing, having a gut feeling, and making assumptions have different meanings. However, they appeared to be used interchangeably. Future work should determine whether logical rationale was present (gut feeling), a proper understanding of the problem (intuition), theory to test against (hypothesis), belief with no proof (assumption) [27,28,44–46].

Moreover, the possibility that confirmation bias and other biases influenced design actions were likely. By stating that some designs are going to be better (whether it is based on factual information or not), displays a participant's tendency to seek and interpret evidence to confirm existing beliefs which aligns with prior work on confirmation bias in design cognition [16,46] and hypothesis testing [45]. This result matters because participants may perform design actions that reinforce their biases or disregard information counter to their approach. Without using concept-selection methods and tools, unexpected design actions and reasons are likely to increase, resulting in low-performance designs outside the lab.

5.3 Lessons for future studies that link design actions, reasoning, and outcomes for an individual's concept-selection process.

Constraining the design space enabled this research to directly compare design actions between participants and reduce subjectivity that an unconstrained design solution space would have introduced. However, this constraint was also a limitation to designer creativity and the transformations possible. While a 10-minute task and five tests provided meaningful results, future work could consider modifying the task duration and number of tests. Self-reported design reasoning via a survey completed remotely was subjective and could be improved. Participants communicated their processes with varying terminology [42,43] and under/over-stated their process. Perhaps a short interview could follow, or a more detailed survey on their process could be utilized. With over 50 participants, this study collected multiple decision elements (actions, rationale, and performance) from a single study in one sitting, which is highly desirable for the design research community [18]. Traditionally multiple studies would need to be carried out to gather the information conducted in this study. Future work will continue to use this constrained design solution space to explore additional influences on participants and how design actions, reasoning, and design outcomes change with the introduction or reduction of new information.

6. CONCLUSION

This paper presents an empirical study combining qualitative and quantitative methods connecting design actions, reasoning, and outcomes, and makes a series of recommendations regarding decision-making strategies for improving design performance. Using a robotic gripper design problem as a test scenario, detailed design process data from 68 participants was collected with a graphical user interface and post-task survey. Results highlight specific design actions that improve design outcomes, such as the use of testing strategies and selecting a final design for which maximum task performance was observed (e.g., rather than guessed). Final design performance, and designer confidence in those values, highly influence a designer's willingness to submit an alternative solution. Although design reasoning (the decision-making heuristics underpinning observed behavior) explains design actions, participants with similar importance ratings for design reasoning factors did not have similar design actions, nor did they achieve similar outcomes. Understanding why designers select lower performance designs can be partially explained by design actions but not by design reasoning.

ACKNOWLEDGMENTS

This work has been supported by the University of California Regents and partially supported by the National Science Foundation under Grant No. DGE-1633740. The findings presented in this work represent the views of the authors and not necessarily those of the sponsors.

REFERENCES

[1] Kudrowitz, B. M., and Wallace, D., 2013, "Assessing the Quality of Ideas from Prolific, Early-Stage Product Ideation," J. Mech. Des, **24**(2), pp. 120–139.

[2] Clark, P. M., and Mirels, H. L., 1970, "Fluency as a Pervasive Element in the Measurement of Creativity," Journal of Educational Measurement, **7**(2), pp. 83–86.

[3] Mirabito, Y., and Goucher-Lambert, K., 2020, "The Role of Idea Fluency and Timing on Highly Innovative Design Concepts," *Proceedings of the IDETC/CIE*, Virtual, Online, August 17–19, 2020, ASME Paper No. DETC2020-22593.

[4] Zheng, X., Ritter, S. C., and Miller, S. R., 2018, "How Concept Selection Tools Impact the Development of Creative Ideas in Engineering Design Education," J. Mech. Des, **140**(5).

[5] Matejka, J., Glueck, M., Bradner, E., Hashemi, A., Grossman, T., and Fitzmaurice, G., 2018, "Dream Lens: Exploration and Visualization of Large-Scale Generative Design Datasets," *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, Montreal QC Canada, pp. 1–12.

[6] Toh, C. A., and Miller, S. R., 2015, "How Engineering Teams Select Design Concepts: A View through the Lens of Creativity," Design Studies, **38**, pp. 111–138.

[7] Kazerounian, K., and Foley, S., 2007, "Barriers to Creativity in Engineering Education: A Study of Instructors and Students Perceptions," J. Mech. Des, **129**(7), pp. 761–768.

[8] Rietzschel, E. F., Nijstad, B. A., and Stroebe, W., 2010, "The Selection of Creative Ideas after Individual Idea Generation: Choosing between Creativity and Impact," British Journal of Psychology, **101**(1), pp. 47–68.

[9] Zheng, X., and Miller, S. R., "Out in the Field Versus Inside in the Lab: A Comparison of Design Professionals' Concept Screening Practices," J. Mech. Des, pp. 1–47.

[10] Ulrich, K., 1995, *Product Design and Development*, McGraw-Hill.

[11] Hay, L., Duffy, A. H. B., McTeague, C., Pidgeon, L. M., Vuletic, T., and Grealy, M., 2017, "Towards a Shared Ontology: A Generic Classification of Cognitive Processes in Conceptual Design," Des. Sci., **3**, p. e7.

[12] Keeney, R., and Raiffa, H., 1993, *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*, Cambridge University Press, New York.

[13] Pahl, G., Beitz, W., Feldhusen, J., and Grote, K.-H., 2007, *Engineering Design: A Systematic Approach*, Springer-Verlag, London.

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

[15] López-Mesa, B., and Bylund, N., 2011, "A Study of the Use of Concept Selection Methods from inside a Company," Res Eng Design, **22**(1), pp. 7–27.

[16] Hallihan, G. M., and Shu, L. H., 2013, "Considering Confirmation Bias in Design and Design Research," Journal of Integrated Design and Process Science, **17**, pp. 19–35.

[17] Goel, V., "Cognitive Processes Involved in Design Problem Solving," *Sketches of Thought*, MIT Press, pp. 95–126.

[18] Hay, L., Duffy, A. H. B., McTeague, C., Pidgeon, L. M., Vuletic, T., and Grealy, M., 2017, "A Systematic Review of Protocol Studies on Conceptual Design Cognition: Design as Search and Exploration," Des. Sci., **3**.

[19] Goldschmidt, G., 2014, *Linkography: Unfolding the Design Process*, MIT Press.

[20] Sonalkar, N., Mabogunje, A., and Leifer, L., 2013, "Developing a Visual Representation to Characterize Moment-to-Moment Concept Generation in Design Teams," International Journal of Design Creativity and Innovation, **1**(2), pp. 93–108.

[21] Pirolli, P., and Card, S., 1995, "Information Foraging in Information Access Environments," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, USA, pp. 51–58.

[22] Berg, B., Stylianou, A. C., and Mezei, R. A., 2018, "Information Foraging—A Model for Exploration Breadth and Depth," Information Systems Management, **35**(2), pp. 161–180.

[23] Papalambros, P. Y., and Wilde, D. J., 2000, *Principles of Optimal Design: Modeling and Computation*, Cambridge University Press.

[24] Oliveira, A., 2007, "A Discussion of Rational and Psychological Decision-Making Theories and Models: The Search for a Cultural-Ethical Decision-Making Model," Electronic Journal of Business Ethics and Organization Studies, **12**(2), p. 6.

[25] Hazelrigg, G. A., 1997, "On Irrationality in Engineering Design," J. Mech. Des, **119**(2), pp. 194–196.

[26] Savage, J. C. D., Moore, C. J., Miles, J. C., and Miles, C., 1998, "The Interaction of Time and Cost Constraints on the Design Process," Design Studies, **19**(2), pp. 217–233.

[27] Galle, P., 1996, "Design Rationalization and the Logic of Design: A Case Study," Design Studies, **17**(3), pp. 253–275.

[28] Galle, P., and Béla Kovács, L., 1996, "Replication Protocol Analysis: A Method for the Study of Real-World Design Thinking," Design Studies, **17**(2), pp. 181–200.

[29] Jansson, D. G., and Smith, S. M., 1991, "Design Fixation," Design Studies, **12**(1), pp. 3–11.

[30] Mueller, J. S., Melwani, S., and Goncalo, J. A., 2012, "The Bias Against Creativity: Why People Desire but Reject Creative Ideas," Psychol Sci, **23**(1), pp. 13–17.

[31] Nelius, T., Doellken, M., Zimmerer, C., and Matthiesen, S., 2020, "The Impact of Confirmation Bias on Reasoning and Visual Attention during Analysis in Engineering Design: An Eye Tracking Study," Design Studies, **71**, p. 100963.

[32] Karim, A. K. M. R., Proulx, M. J., and Likova, L. T., 2016, "Anticlockwise or Clockwise? A Dynamic Perception-Action-Laterality Model for Directionality Bias in

Visuospatial Functioning," Neuroscience & Biobehavioral Reviews, **68**, pp. 669–693.
[33] Li, M. S., Melville, D., Chung, E., and Stuart, H. S., 2020, "Milliscale Features Increase Friction of Soft Skin in Lubricated Contact," IEEE Robotics and Automation Letters, **5**(3), pp. 4781–4787.
[34] Tabeau, K., Gemser, G., Hultink, E. J., and Wijnberg, N. M., 2017, "Exploration and Exploitation Activities for Design Innovation," Journal of Marketing Management, **33**(3–4), pp. 203–225.
[35] Tahera, K., Wynn, D. C., Earl, C., and Eckert, C. M., 2019, "Testing in the Incremental Design and Development of Complex Products," Res Eng Design, **30**(2), pp. 291–316.
[36] Ngoon, T. J., Walker, C. M., and Klemmer, S., 2019, "The Dark Side of Satisficing: Setting the Temperature of Creative Thinking," *Proceedings of the 2019 on Creativity and Cognition*, Association for Computing Machinery, New York, NY, USA, pp. 591–596.
[37] Parkinson, B., Meacock, R., Sutton, M., Fichera, E., Mills, N., Shorter, G. W., Treweek, S., Harman, N. L., Brown, R. C. H., Gillies, K., and Bower, P., 2019, "Designing and Using Incentives to Support Recruitment and Retention in Clinical Trials: A Scoping Review and a Checklist for Design," Trials, **20**(1), p. 624.
[38] Milgram, R. M., and Rabkin, L., 1980, "Developmental Test of Mednick's Associative Hierarchies of Original Thinking," Developmental Psychology, **16**(2), pp. 157–158.
[39] Verónica, J. R., 2018, *Analyzing the Role of Cognitive Biases in the Decision-Making Process*, IGI Global.
[40] Goldschmidt, G., and Tatsa, D., 2005, "How Good Are Good Ideas? Correlates of Design Creativity," Design Studies, **26**(6), pp. 593–611.
[41] Goucher-Lambert, K., and Cagan, J., 2017, "Using Crowdsourcing to Provide Analogies for Designer Ideation in a Cognitive Study," *21st International Conference on Engineering Design (ICED 17)*, Vancouver, Canada, pp. 529–538.
[42] Krishnakumar, S., Berdanier, C., McComb, C., and Menold, J., 2021, "Lost in Translation: Examining the Complex Relationship Between Prototyping and Communication," J. Mech. Des, **143**(091402).
[43] Rescher, N., 1998, "The Role of Rhetoric in Rational Argumentation," Argumentation, **12**(2), pp. 315–323.
[44] Taura, T., and Nagai, Y., 2017, "Creativity in Innovation Design: The Roles of Intuition, Synthesis, and Hypothesis," International Journal of Design Creativity and Innovation, **5**(3–4), pp. 131–148.
[45] Beattie, J., and Baron, J., 1988, "Confirmation and Matching Biases in Hypothesis Testing," The Quarterly Journal of Experimental Psychology Section A, **40**(2), pp. 269–297.
[46] Dong, A., Lovallo, D., and Mounarath, R., 2015, "The Effect of Abductive Reasoning on Concept Selection Decisions," Design Studies, **37**, pp. 37–58.