Research in Engineering Design (2025) 36:9
https://doi.org/10.1007/s00163-025-00451-3

ORIGINAL PAPER

The effect of sketch and render quality and design experience on concept evaluation in engineering design

Nisha Detchprohm[1] · **Anastasia Schauer**[2] · **Haley Stokes**[1] · **Katherine Fu**[3]

Received: 30 June 2024 / Revised: 16 March 2025 / Accepted: 2 April 2025

Abstract

Illusory correlation (IC) is a cognitive bias that appears when decisions are based on false perception of patterns from limited data and can prevent subjects from detecting present correlations. It appears in design or psychology studies as a secondary bias, but has not been studied in-depth in engineering design. This research examines the presence of IC during concept evaluation of the engineering design process and how current engineering design education may mitigate the appearance of IC. To examine IC, different products at four different sketch quality levels and render quality levels, from quick hand drawing to shaded product render, were presented to participants through a survey-based data collection instrument. The four sketch and/or render quality levels simulate the variety of presented drawings when a design engineer is evaluating colleagues' concepts during group work. Participants, 70 undergraduate (novice) and 21 graduate (advanced) engineering students at a major southeast US institution, were asked to rank these products based on a series of function-based and preference-based attributes. The collected data were analyzed to see if sketch and/or render quality impacted participants' ability to gage functionality of the presented products. Results indicated no statistically significant linear correlation between better rankings of products and the sketch and/or render quality level of the provided depictions for function-based questions; however, a non-linear relationship was present for preference-based questions where participants gave higher rankings to products drawn at intermediate quality levels. No statistically significant differences were found in the strength of correlations between rankings and sketch and/or render quality levels in the comparison of novice and advanced student designers, but advanced student designers' perception of product functionality was more strongly correlated to pre-determined baseline answers based on user ratings of the selected products. This may indicate less vulnerability to IC bias with more design engineering education due to a stronger intuition for and understanding of product functionality based on visual inspection.

Keywords Illusory correlation · Cognitive bias · Concept evaluation · Design expertise · Sketch quality · Render quality

1 Introduction

1.1 Motivation

The current landscape of engineering design work involves the adaptation of ideas to maintain, if not improve, functionality for an intended solution. With the inherent desire for innovation in engineering, it is important to observe whether cognitive biases, deviation from rational or objective thinking (Blanco 2017), significantly impact the quality of concept design evaluations such that better solutions are disregarded. Many types of cognitive bias have been documented extensively in the field of psychology, but research in cognitive bias in engineering design has focused heavily on confirmation bias and sunk cost bias. Sunk cost bias, or the pursuit of a direction due to prior

✉ Katherine Fu
kfu26@wisc.edu

Nisha Detchprohm
ndetchprohm3@gatech.edu

Anastasia Schauer
schauer@utexas.edu

Haley Stokes
haley.stokes17@gmail.com

[1] Georgia Institute of Technology, George W. Woodruff School of Mechanical Engineering, Atlanta, GA, USA

[2] Walker Department of Mechanical Engineering, University of Texas at Austin, Austin, TX, USA

[3] Department of Mechanical Engineering, University of Wisconsin-Madison, Madison, WI, USA

Published online: 23 April 2025

investment despite the risk of further losses, can manifest as design fixation, which can result in the inclusion of common design features and the exclusion of novel features (Koh and De Lessio 2018). Design fixation relates to the sunk cost associated with the creation of physical prototypes and its correlation to the sunk cost of resources and time (Viswanathan and Linsey 2013). Another well-known type of cognitive bias is confirmation bias, which is the tendency to seek or interpret evidence in a way that confirms pre-existing beliefs, reducing effective reasoning (Koriat et al. 1980; Nickerson 1998; Hallihan and Shu 2013; Hallihan et al. 2013). This includes selecting data that meets a pre-determined hypothesis even though the overall data points in a different trend. In a study about cognitive bias, Macomber and Yang found that better, or higher, sketching and rendering quality led to greater preference for those design concepts despite the content of the sketches or renders (Macomber and Yang 2012). Although they did not mention illusory correlation by name, this study confirms the plausibility for illusory correlation, false perception of correlation from provided data, impacting perceived product attributes in the engineering design field.

Illusory correlation is a type of cognitive bias that occurs when conclusions are drawn from incorrect patterns through observations between two distinctive stimuli. Studies in psychology show that these incorrect conclusions can become ingrained in thought and deter correlations that are present, or strengthen a correlation, leading to an overestimation of their co-occurrence, such as in stereotypes. Hamilton and Gifford found illusory correlation to be the foundational thought process behind stereotypic beliefs (1976). In their experiment, subjects associated undesirable or desirable behaviors with members of an imaginary minority group in accordance with which behavioral trait occurred less frequently. They observed that subjects were likely to form strong illusory correlations between two infrequently occurring stimuli, as the distinctiveness of these pairs allowed for easier information recall.

The presence of illusory correlations also pervades historical and modern sociological developments. Deininger et al. (2019) explored a part of this perception, looking into how customer perception of early design concepts change based on prototype type and quality, along with other context-specific factors like culture. For an everyday example, when heavy lifting is required for a task, men are often selected first. Sometimes this request is asked with a rendition of the phrase "looking for two strong gentlemen," which eliminates other genders from consideration. This can be problematic combined with the lack of diversity in the technology sector, resulting in illusory correlation bias during the creation of these technologies. An example in modern technology is the disproportionate misidentification of minorities with facial recognition. Without diversity, products are created that only adequately service people with similar traits to their creators, creating a false correlation of consistent functionality because the observations from testing prove functionality within those bounds.

1.2 Research questions

The presence of stereotypic beliefs from illusory correlations translates into biased assessments of products during the engineering design process, and results in non-inclusive and potentially ineffective product design. Two studies analyzed this and inspired the research question through their findings and survey set-up.

Within engineering design, Fillingim examined the presence and the manifestation of cognitive bias using survey-based data collection to gage student perception on their design concepts (Fillingim et al. 2023). They implemented five surveys across a semester long project at different points of the engineering design process. The results showed that subjects portrayed common cognitive biases with their selected team project, confirming the plausibility of focusing on the appearance of a specific bias during the concept evaluation process.

Macomber and Yang examined partiality in sketch finish and style using three objects of varying complexity that were compared separately. There were four set line drawing levels and four set shaded drawing levels that ranged from hand drawings to CAD models. Each object had one selected image that was drawn eight times, once for each set drawing level. They tested for partiality in the line drawings and then shaded drawings by asking about perceived drawing attributes and subjective opinions within a ranking survey. The results showed that subjects preferred realistic, clean hand drawings to rougher drawings or CAD models (Macomber and Yang 2012).

This work looks to expand the two aforementioned studies by exploring the overarching research question: *how does illusory correlation impact engineering designers' perception of functionality of products during concept evaluation?*

By addressing this question, future recommendations could be made regarding how concepts should be presented to remove as much illusory correlation as possible. A variety of products within the same product groups expand the study beyond objects, and adding function-based question adds relative ranking between different products within the same product group. Comparing novices to advanced student designers will determine if the education level and/or amount of design experience has an impact on the appearance of illusory correlation. Changes to engineering design curriculum could be recommended if there is a noticeable difference between novice and advanced student designers. The significance in these findings could serve to motivate interventions to mitigate illusory correlation within the engineering design process.

2 Literature review

2.1 Illusory correlation bias

Illusory correlation bias appears when choices are made based on false perception of patterns or correlations from limited data (Gilovich et al. 2002; Hallihan et al. 2013). It is a cognitive bias that often appears in design- or psychology-related studies as a secondary bias; however, it is not heavily focused on if mentioned at all. An observational report to learn more about the factors of illusory correlation states that the bias appears as a systematic error from humans trying to predict relationships between multiple stimuli, specifically when those two events are: “not correlated, correlated less than reported, or correlated in the opposite direction” (Chapman 1967). These findings focus on the strong association between words and distinctive stimuli, which will be discussed further in the psychology portion of this literature review.

Following the work of Chapman, Tversky and Kahneman (1974) enhanced the definition of illusory correlation bias. They stated that it stems from the availability heuristic due to the judgment of paired items by the frequency of their co-occurrences, repetition, natural associations, pair-distinctiveness, and overall strength between associative bonds. It was determined that illusory correlation bias is resistant to the presentation of contradictory data, such that it persists even when a negative correlation is established between associated items (Tversky and Kahneman 1974). Strong association between items leads to the conclusion that they were frequently paired in prior experiences, and many psychological studies utilize this fact as a part of their methodology.

Most notably, Hamilton et al. (1985) applied this information to show how illusory correlation affects information recall and pattern recognition abilities. Within their study, they found that illusory correlation develops during the encoding of a stimuli sequence of distinctive, co-occurring events, and that subjects affected by illusory correlations were able to recall a higher proportion of distinctive, co-occurring stimulus events. Additionally, when subjects were asked to judge how much they liked each stimuli pairing, subjects developed much stronger opinions of the distinctive, co-occurring events for which illusory correlation biases were formed. These effects on information recall were attributed to the differential processing of the infrequently occurring stimulus items. As an example, subjects could be shown a series of shapes with different colors as a stimuli sequence, where each shape and color combination occurs at the same rate. A distinctive, co-occurring event for this study could be a red octagon since subjects may more easily recall this shape due to its association with stop signs. Alternatively, this study may show different color and shape combinations at different rates such that a blue circle occurs frequently and a yellow square is rarely shown. According to these findings, subjects would be better able to recall the occurrence of the yellow square since this shape and color combination is distinct in its infrequency and thus subjects would process these stimuli differently by identifying its rarity. A similar study found that serial presentation of information resulted in an illusory correlation such that participants judged one event to be contingent on the other (Ward and Jenkins 1965). These works provide a basis for the review of literature related to illusory correlation discussed in the following sections, particularly in the field of psychology, although this bias is a worthy research topic for a variety of areas of application and fields of study.

2.2 Illusory correlation in psychology research

Illusory correlation has most comprehensively been observed within the field of psychology through a variety of studies that aim to form the bias within subjects and observe its unique effects on information recall and judgment (Chapman and Chapman 1967; Tversky and Kahneman 1974; Hamilton et al. 1985). Chapman and Chapman (1967) observed the formation of illusory correlation bias during clinical diagnosis by psychologists and psychiatrists. Through six experiments that paired drawings of a potential patients and statements regarding their symptoms, it was found that illusory correlation bias formed and persisted “both under repeated exposure to the stimulus materials and under conditions designed to maximize both motivation and opportunity to observe [the patient] accurately.” A notable experiment from this study showed the subjects the drawings of patients with two symptoms each such that the symptoms were randomly distributed to the drawings and no correlation between physical appearance and symptoms of the patient drawings existed. Despite this, subjects developed an illusory correlation between the two stimuli which would result in incorrect psychodiagnosis of a patient. These correlations included associating concerns about intelligence with having an emphasized head, concerns with masculinity with muscular figures, suspiciousness with atypical eye drawings, and many more. It was also observed that the illusory correlations developed did not attenuate when subjects were presented with information that disproves their assumptions, even when the subjects were motivated to make accurate patient observations and could view the stimulus materials for unlimited amounts of time. Therefore, illusory correlation bias in psychodiagnostics could be mitigated by subjecting graduate psychology students to trainings, such that the students may become aware of their biases (Chapman and Chapman 1967).

Hartsough (1975) followed this work by observing how the strength of associations between stimuli effects the strength of the formed illusory correlation bias. This study presented words with colors such that every color had two associated words and each set of words and colors had a different associative strength. For example, the set with the greatest association factor was "red, blood, hurt" and the set with the weakest association was "yellow, butter, bread." Each color was paired with one word and placed onto a card for subject viewing such that each word–color combination only occurred once out of 40 times. When asked to access the frequency of word–color combinations from memory, the associative connections between stimuli resulted in illusory correlation bias formation. This was indicated by an increase in illusory correlation strength accompanying a decrease in strength of a pair's associative connections (Hartsough 1975), supporting findings that increasingly distinctive stimuli are most easily recalled by subjects (Hamilton et al. 1985).

Illusory correlation bias has also been studied as it emerges during psychodiagnosis across undergraduate and graduate psychology students (Starr and Katkin 1969). This study design required subjects to pair a list of potential patient symptoms with fill-in-the-blank sentences that were completed by the patients. An example from the study is the pairing of, "He complains of perpetual fatigue and illness," as a patient symptom with the fill-in-the-blank sentence being "At bedtime — I cry myself to sleep" (Gilovich et al. 2002). Observations regarding the formation of illusory correlation confirmed the foundational observations of Chapman and Chapman (1967), including its reliance on associative connection and stimuli pair frequency. Subjects from all education levels were also asked to rate their confidence in patient observation as "positive, fairly sure, or guessing." The results of these ratings were that graduate students "overlearned the notion that there are few unequivocal psychological relationships," which made them less likely to be certain of their observations (Starr and Katkin 1969). These findings indicate that with increased education level, an individual will form weaker illusory correlations.

A critical area of illusory correlation research that has not yet been discussed are studies that reject the hypotheses of foundational illusory correlation studies. Most notably, Fiedler et al. (1993) found that distinctiveness does not improve an individual's ability to recall that information, but rather, infrequent observations result in information loss or impaired memory. These findings were presented as a rejection of Hamilton and Gifford's "paired-distinctiveness account" (1976) through both subject experimentation and computerized simulations of illusory correlation development in humans. The failure of the paired-distinctiveness account was attributed to a lack of attempts "made to manipulate or control distinctiveness," as well as the perceived ability to recall infrequent information being due to "correlational and often inappropriate analysis," leading researchers to conclude that "illusory correlations may reflect differential information loss for small versus large [stimuli] groups" such that "information processing is impaired" (Fiedler et al. 1993).

Given the multitude of proposed factors that may influence the formation of illusory correlation and the abundant evidence of problems it can create across disciplines, investigation of the bias in the critical early stages of the engineering design process is warranted. Illusory correlations throughout the engineering design process are the basis of some of the problems people face in day-to-day life. For example, car crash test dummies have historically been designed as an "average-sized male body" despite this profile of a person sustaining "the fewest injuries in automobile accidents" (Linder and Svensson 2019). The automotive industry falsely correlates safety of their test dummies to the safety of all possible users is reflective of an illusory correlation bias problem (Cham and Yang 2008) as the conclusion from testing neglects to include the variety of people who will eventually drive the tested cars. The present study will assist in identifying some key factors that determine the formation and intensity of illusory correlation such that the engineering design community can learn how to create better and more inclusive products.

2.3 Illusory correlation in engineering design research

As contributing to engineering design research is the primary objective of this work, literature was examined to determine how design thinking is impacted by varying levels of sketch and render quality. When designers create concept sketches or CAD renderings themselves, it has been found that there is no correlation between the value of the design outcome and the sketching skill level (Cham and Yang 2008). Despite this, user response to these sketches and renders throughout the early stages of design is highly influenced by the level of skill and quality presented as found by Macomber and Yang (2012). In combining these two findings, a new question emerges in regard to a designer's ability to assess design concept quality and end-user satisfaction from sketches and renderings that are not their own.

This is a critical area of research as engineering design processes inherently require group work for successful early-stage concept evaluation. The emergence of illusory correlations during this process can result in the creation of products that do not adequately meet the needs of the user. Examining the role of education and relevant experience in the formation of illusory correlations is significant since this bias stems from the availability heuristic, which influences memory recall through an event pair's frequency of co-occurrence, natural association, and repetition (Tversky and Kahneman 1973). For engineering design, it is unclear whether more design education and experience mitigates or worsens the development and strength of illusory correlation bias.

Engineering design expertise can result in a number of logical fallacies and cognitive biases, including the formation of illusory correlation biases that are highly resistant to contradictory data (Strigini 1996). From their own individual design experiences, these correlations could become intuitive design thoughts and decision-making tools regarding the construction and assessment of sketches and renders. Additionally, overconfidence or self-affirmation has the potential to strengthen illusory correlations when experts are presented with contradictory information (Strigini 1996; Munro and Stansbury 2009). Confirmation bias has the potential to emerge with relative ease such that, during early concept evaluation, only one configuration is fixated on and iterated upon (Cross 2004). Despite the potential for fixation to emerge, experienced designers who acted intuitively in design tasks commonly created a positive design outcome (Cross 2001).

Häggman et al. (2015) examined user preferences in early-stage engineering design by observing how subjects rated sketches, prototypes, and computer-aided design models of a product on their perceived attributes and qualities as a function of how long each method took to produce a design. Subjects rated the three design methods on usefulness, creativity, comfort, likelihood of buying, esthetics, clarity of design, and how good the overall design was. This study found that foam prototypes were generated faster than other design methods and resulted in better perceived comfort, creativity, and esthetics. The most relevant finding from this study to this research is that "a novel form alone is sometimes not sufficient for a well perceived design" (Häggman et al. 2015). Novel designs received good creativity ratings from the users but did not score well in other categories due to unfamiliarity, indicating that subjects may produce the illusory correlation that novel products are less functional or reasonable designs. A separate study validated this inverse relationship between perceptions of novelty and functionality; designs that were perceived as esthetically appealing were also more likely to be perceived as novel and surprising, but less functional and useful (Han et al. 2021).

In a study with "naïve", "novice", and "expert" designers, Crismond (2001) investigated how education level affects a designer's ability to thoroughly analyze and redesign simple mechanical devices. The findings of this study echo previous studies regarding the conclusion between the sampling of undergraduate and graduate psychology students, such that in the engineering design discipline, expert subjects "sought to learn what the critical design problem was for each device" and were able to "apply abstract concepts" to the products. Naïve and novice student designers comparatively did not think as critically of the products as they considered redesign options although being prompted by questioning helped the designers understand how to incorporate science skills and ideas into their redesigns (Crismond 2001). This work serves to further distinguish how increased education level may affect the formation of illusory correlation bias during the engineering design process.

Although not often discussed in engineering design literature, the potential for the formation of illusory correlation throughout the design process is evident, and this work aims to inform the design community of the pathways on which this bias may form.

2.4 Conclusions from literature: hypotheses

From this literature review, two more specific research questions and, consequentially, two hypotheses have been formed to address the presence of illusory correlation during concept design and evaluation. The research questions are as follows:

RQ1. How does render and sketch impact a designer's perception of the functionality of a design concept?

It is believed that illusory correlation will cause designers to perceive designs as more functional and exhibit a higher preference for designs that are presented with higher, or better, sketch and render quality (H1) since it has been shown that design stakeholders have higher preference for higher-quality sketches and renderings (Macomber and Yang 2012).

RQ2. How does level of expertise affect the presence or intensity of illusory correlation for designers?

Advanced student designers are expected to develop weaker illusory correlations as their design knowledge and training is anticipated to mitigate or eliminate the bias (Murphy et al. 2011). The effects of education level on the appearance of illusory correlation have primarily been studied for medical disciplines (Jenkins and Ward 1965) although results do not definitively indicate that advanced education levels mitigate the bias. To address this knowledge gap, the impacts of design experience level regarding education on the presence and intensity of illusory correlation will be examined for engineering design students. Novice student designers and students with less design experience are expected to develop more frequent and intense illusory correlations (H2) since they lack the expertise to comprehensively evaluate a design concept (Cross et al. 1994).

3 Methodology

3.1 Overview of the study

The purpose of this study is to monitor the occurrence of illusory correlation bias by observing the effects of sketch and render quality level on perceived product functionality. Using three product groups with four products in each group of varying complexity to simulate the concept evaluation process, this study translates the measurement of illusory correlation from the psychology literature into the engineering field. This study explores the prior work by Macomber and Yang (2012), examining perceived functionality rather than user preference, as modulated by sketch and render quality level.

Three product groups were chosen for this study, coffee makers, washers, and scissors. Cylinders were chosen as a fourth object that is devoid of functionality for the purpose of comparison to the products. The chosen product groups are ones that participants were presumed to be familiar with; however, the selected products within each group were expected to be of varying familiarity due to uniqueness. Four individual products were chosen within each product group as shown in Table 1.

Each product was depicted in eight different ways: with four different sketch quality levels and four different render quality levels; example depictions with full descriptions of these levels can be seen in Table 2. The descriptions were aligned with student training on sketches in design engineering courses during team evaluation. At minimum, the product is clearly outlined to distinguish from set-up

Table 1. Product groups and individual products, depicted at sketch IV

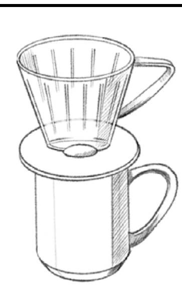
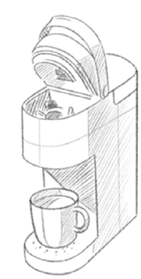
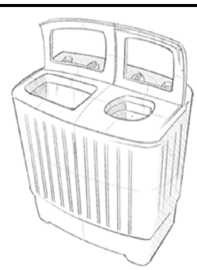
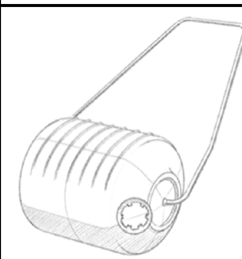
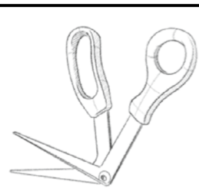
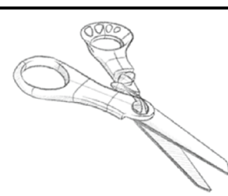
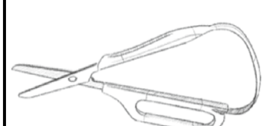
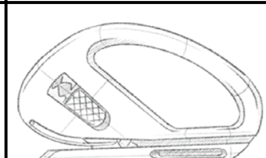
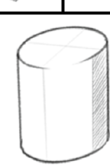
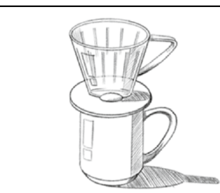
Coffee Makers	Pourover	French Press	Keurig Cup	Moka Pot
				
Washers	Portable Washer Dryer	Washboard and Basin	Pedaled-power Washer	Hippo Roller
				
Scissors	Right Angle	Spring Loaded	Easi-grip	Zippy Cutter
				
Cylinders				

Table 2. Descriptions of sketch and render quality levels with the pourover from the coffee product group as an example depiction for reference

	Level		Description
Sketch Quality Levels	SI		Sketch looks like the beginning of a product drawing that focuses on the overall shape. This will serve as your underlay for all sketch and render quality levels. Product edges are multiple lines and have all of their construction lines using the pencil tool. Use similar line thickness across the entire sketch.
	SII		Sketch with construction lines that define edges and includes general placement of small details. Some or all of the edges can have multiple overlapping lines as long as it illustrates a distinct edge. Contour lines that show geometry should be included. It should be the previous sketch quality level (SI) with less construction lines and clearer edges.
	SIII		Sketch containing clean lines with varying thickness. No construction lines should be visible. All lines should be distinctly singular and faint contour lines should be included where necessary. Light shading can be used in addition to the geometry lines to better visualize the shape. Small details should be included with their respective 3-dimensionality in addition to their relative placement on the product, if it is visible. For example, any buttons or knobs should have a visible thickness in the sketch. It should be the previous sketch quality level (SII) with cleaner edges, no construction lines and additional details.
	SIV		Sketch with clean, singular lines with varying thicknesses to show more detail. Contour lines should be included where necessary, and light hatching can be used to emphasize curvature. Small details should be included and well defined. It should be the previous sketch quality level (SIII) with larger variation in line thickness, more defined details and reflective of the material texture.
Render Quality Levels	RI		Hand drawn sketches with clear edges and light shading with the drawing tool to better show the geometry. The lines within the drawings should use the pen tool starting at this stage. Light yet distinct lines can be added for the geometry wherever necessary. Use either one or two levels of shading that is hand-drawn to increase 3-dimensionality. No cast shadows should be used in these drawings. It should look like the fourth level of sketch (SIV) with defined hatching as the method of shading.
	RII		Hand-drawn rendering with gradients in shading to better visualize organic shapes. Cast shadows are a rough outline of the shape with light hatching to fill it. This render will be the previous with more variation in shading. This render should look like the previous level (RI); however, the hatching should be finer and varying in shade to look more realistic.
	RIII		Rendering with markers in addition to the hand drawing and shading. The levels should be similar to the provided examples and consistent across all the product groups. This should look like the previous render quality level (RII) with grayscale markers or software added shading to look more realistic.
	RIV		This render should only have shading with grayscale markers. Include a cast shadow and add lines that reflect the material of the product. Construction lines should not be visible. It should look like the previous render quality level (RIV) with the hatching replaced by grayscale marker shading. This final drawing should look like a grayscale final product presentation.

lines (SI) and follow steps to clean up concepts to a fully rendered, shaded, and material assigned product (RIV). The individual descriptions are given to the artist to guide their sketching and rendering for the creation of the study materials with all the definitions and resulting depictions in Appendix 1.

3.2 Participants

Study participants were divided into two groups: novice student designers and advanced student designers, referred to as novice and advanced for brevity. Novice student designers were undergraduate mechanical engineering students taking an introduction to engineering graphics and visualization course. Among the 70 novice student designers that agreed to participate, there were 53 men, 16 women and 1 preferred not to say gender, with an average age of 19.3 years old. Race collected from novice participants showed that 24 identified as White / Caucasian, 29 as Asian or Pacific Islander, 7 as Hispanic / Latino/a/x, 4 as African American, 4 as "other", and 2 preferred not to report their race. Advanced student designers were graduate mechanical engineering students taking a course on designing open engineering systems. Among the 21 advanced student designers that agreed to participate, there were 16 men and 5 women, with an average age of 23.2 years old. Race collected from advanced participants showed that 8 identified as White / Caucasian, 12 as Asian or Pacific Islander, and 1 as Hispanic / Latino/a/x.

An important assumption to confirm was that novice student designers had less engineering design exposure as it was their first or second year within their engineering education, compared to the advanced, who were graduate students and had taken design classes for their undergraduate degrees. Academic design experience data confirmed advanced student designers had more exposure to design with an average of 2.8 design classes taken, whereas novice student designers had an average of 1.1 design classes.

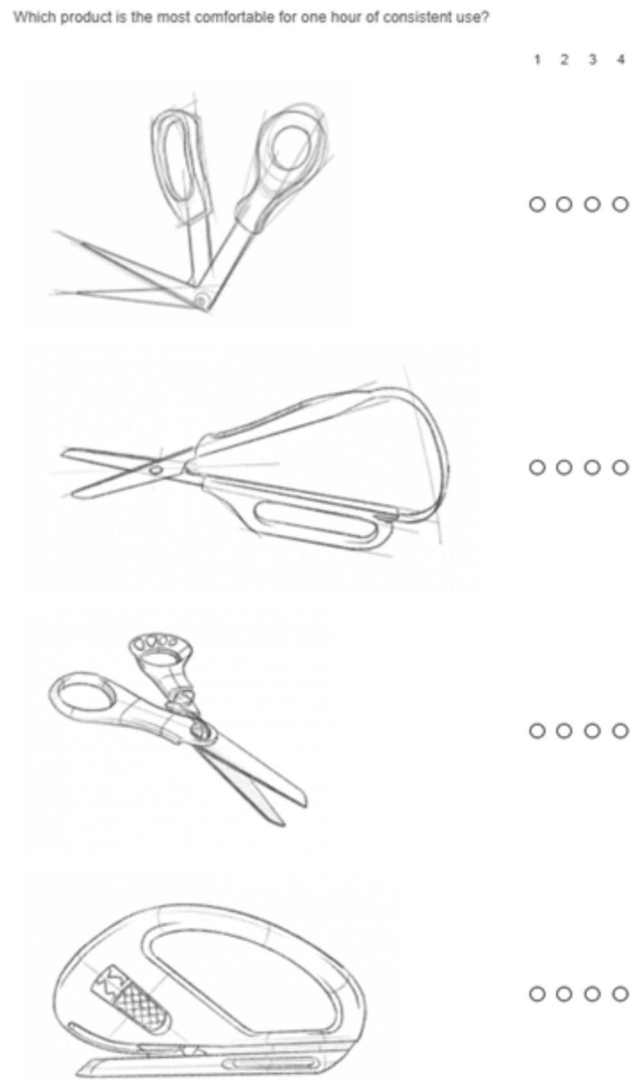


Fig. 1 Survey question for scissor product group. In order from top to bottom: product 1 – sketch quality level I, product 2 – sketch quality level II, product 3 – sketch quality level IV, product 4, sketch quality level III. Participants provide their ranking from 1 (best) to 4 (worst) using the radio buttons to the right

3.3 Survey design

To measure the effect of sketch and render quality level on the appearance of illusory correlation, a survey was designed. The survey required participants to make decisions in a fast-paced manner by ranking products of varying sketch and render quality level relative to each other. Radio buttons were placed to the right of the drawings and were used to rank the choices with 1 as the best, 4 as the worst based on the relevant attribute. A sample question is shown in Fig. 1. Each question prompt and instructions for ranking appeared on its own page with the following page showing the same question with the answer choice options revealed for ranking and a timer to record the total time the participant spends on the page. This accounted for the variability of reading speeds across participants while keeping the measurement method for selected answer time consistent.

As described in Table 2, there were four sketch quality levels and four render quality levels tested in this study. The survey was divided into two equal halves: first, testing the effect of only sketch quality level, referenced as sketch depictions, and second, testing the effect of only render quality level, referenced as render depictions, with no overlap between the two. In each question of the survey, each sketch or render quality level appeared once across all four ranking options. For example, in Fig. 1, sketch quality levels SI, SII, SIII, and SIV are each shown, each paired with a different scissor product. Within each half of the survey, there were four sections, one for each product group (three total) and one for cylinders. The sections for the three product groups had eight functionality questions and two personal preference questions, and the cylinder section had four personal preference questions, totaling 34 questions in one-half and 68 questions in the full survey. The flowchart showing the order of questions within and across product groups for the sketch quality levels, and the survey questions used to collect functionality and preference data can be found in Appendix 2.

The participants were told to complete their rankings in 12 s or less to prevent other biases from appearing. This timing factor is based upon prior work by Dane and Pratt (2009), who used the "direct intrusion method" in studying design problem-solving, during which a participant is given a design problem and must immediately offer a solution. In a different study, Dane et al. (2012) gave participants a short period of time of 5 s, called "immediacy", to think about a problem intuitively compared to a longer period of 30 s to think analytically, demonstrating that the period of time given to designers for decision-making can vary their mode of thinking. Participants are not provided a timer to avoid creating a false sense of urgency and the analysis of the results accounts for time variation by removing any rankings that take longer than 30 s.

In order to measure whether participants were providing biased responses to the functionality survey questions as influenced by sketch or render quality level, baseline answers had to be chosen, against which the survey answers could be compared. Previous studies used online user reviews from those who purchased the products as baseline answers because of the impact of user perception on new product development (Li S. et al. 2021), (Jiang et al. 2017). Additionally, the comparison of user reviews to designer perception tracks whether designers can gage how users perceive attributes of a design concept (El Dehaibi, N. et al. 2019). For this study, these are referred to as baseline answers and chosen by examining product specifications across multiple sources for quantitative characteristics and finding the average answers. For less quantitative characteristics, such as comfort of use, customer, or user, star ratings compiled from a major online retailer. The findings are compiled into an evaluation matrix to distinguish the rankings between products. Attribute questions for function-based and preference-based questions, listed in Table 7, were selected based on previous studies from the literature (Häggman et al. 2015) and most frequently asked questions found when searching for baseline answers.

For one question, there are 24 possible combinations for sketch and/or render quality level assignments of individual products without repetition of level in the four answer choices. The survey consisted of eight sections, two for each of the three product groups and two for cylinder groups, one sketch and one render depiction each. By varying sketch and/or render quality level assignments, each combination was distributed as evenly as possible among the products and participants, with no repetition of combinations in an individual section. A randomly ordered list of numbers 1 through 24 was used to determine how the combination assignments were distributed in each section across each survey version.

3.4 Study procedure

Both groups received a ten-minute briefing on the study procedure and an online waiver of written consent information. Study and consent procedures were conducted under the guidance of the institutional review board at Georgia Institute of Technology. If participants chose to proceed, they indicated informed consent and continued the questionnaire. Participants who did not wish to give consent and participate in the study exited the online questionnaire at that time. Participants agreed to take the survey with compensation of extra credit for their respective course. For those who chose not to participate, an alternative extra credit assignment was offered. Participants then answered a few short background questions regarding their demographic information and prior design experience.

Participants then responded to all the study questions, grouped by product type and sketch and/or render quality level, totaling eight sections. The order in which the grouped product type sections appeared was randomized by the survey software and the cylinder sections appeared last for all the surveys. The functionality questions within each product group section were randomized by the survey software. The order of survey questions is visualized in Fig. 13. There was a timer on the ranking questions to record the time participants took to answer. The survey continued in the same format until all eight sections were complete, taking approximately 45 min total.

3.5 Assessing the effects of sketch and render quality level

The data were analyzed through multiple coding schemes in SPSS to measure the illusory correlation bias between a selected answer choice and sketch or render quality level. Analysis was separated between functionality-based questions and preference-based questions, where the functionality-based questions were compared to the baseline answers. Illusory correlation bias was identified by comparing correlations between ranking choices and sketch or render quality level to the correlation between ranking choices and the baseline answer. Data were further analyzed by comparing design expertise level.

4 Results

The results of the study are presented as structured by the two research questions. The analysis was completed with one filter on the dataset. Survey responses that were completed in over 30 s were removed from the analysis due to the impacts on time-constrained decision-making (Baylor 1997; Dane and Pratt 2009; Dane et al. 2012). It should also be noted that all incomplete responses (questions whose answers were missing more than one ranking) were not removed from the data set prior to analysis because the selected analysis methods can still incorporate the given responses within the incomplete data sets.

After filtering the results, Kendall's tau-b is used for both research questions and hypotheses to compare two sets of rankings in order to determine the strength and direction of correlation between them. The outcome of the Kendall's tau-b test is a correlation coefficient (τ_b) and an associated significance (*p*). For all analysis, a p value at or below 0.05 is considered statistically significant. Values of τ_b in terms of strength of correlation are as follows with weak strengths considered sufficient enough to support future studies (Botsch 2011):

- Less than + or − 0.10: very weak
- + or − 0.10 to 0.19: weak
- + or − 0.20 to 0.29: moderate
- + or − 0.30 or above: strong

Since the participants rank best as 1 and worst as 4, but the sketch or render quality levels go from the lowest quality as 1 and the best quality as 4, the correlation coefficients for all data sets grouping functionality and preference-based questions are negated to present results as a correlation between increased functionality and better depiction quality. For comparisons between functionality-based questions and baseline answers, the correlation coefficient is not negated because the participant responses and the baseline answers follow the best as 1 and the worst as 4.

The filtered responses used and other statistical tests will be stated under their corresponding research question.

4.1 Research question 1: correlating functionality with sketch and render quality levels

The first research question focused on testing if participants, regardless of their level of expertise, exhibited illusory correlation in their evaluation of the products. The first hypothesis states that illusory correlation will cause designers to perceive designs with higher, or better, sketch and render quality as more functional, resulting in better ratings for these designs (H1). For all selected data sets in this section, a significant correlation between product functionality characteristics and sketch and/or render quality level indicates the manifestation of illusory correlation.

The first section looks for the presence of illusory correlation in the product groups for sketch and/or render quality levels with both function-based and preference-based questions are combined. The cylinder group is used to confirm that the survey design validates previous results from Macomber and Yang and serves as a good baseline to analyze product groups. If there is a significant correlation, then H1 is supported because participant responses show that provided sketch and/or quality level impacted ranking. Based on the Macomber and Yang study, the second section covers an additional analysis to see if there was a linear correlation between better sketch or render quality levels and participant rankings for product groups.

The third section looks for the presence of illusory correlation in the perceived functionality of product groups for sketch and/or render quality levels. All participant rankings for the function-based questions only for the three product groups are combined and compared to the baseline answers. The cylinder group was not included because it does not have function-based questions. If there is a significant correlation between participant rankings and baseline answers, then H1 is not supported because participants are able to perceive functionality irrespective of the shown sketch and/or render quality level.

4.1.1 Correlations between participant rankings and sketch and render quality levels using Kendall's Tau-b

The data for all participants, across both levels of expertise, for all product groups (coffee makers, washers, and scissors) and function and preference-based questions were combined; then the sketch quality levels only, render quality levels only, and both sketch and render quality levels combined were compared to one another. Cylinders were not included in the compilation because their set of preference-based questions differed from the rest of the products, as seen in Table 7. Kendall's Tau-b was used to determine if higher rankings were associated with higher sketch and/or render quality levels of product depictions.

There were non-significant negative correlations between sketch and/or render quality levels and participants' ranking of all product groups based on sketch quality only ($\tau = -0.012$, $p = 0.145$), render quality only ($\tau = -0.008$, $p = 0.319$), and combined quality ($\tau = -0.01$, $p = 0.083$). The lack of statistically significant correlations between participant judgment of product function and sketch and/or render quality level indicates that Hypothesis 1 was not supported, and that illusory correlation did not significantly impact participants' perceptions of the products.

Cylinders group data following the same grouping, all participants and all questions (preference-based only for cylinders), were combined and compared across sketch quality levels only, render quality levels only, and both sketch and render quality levels as a benchmark devoid of function. The results were also included to benchmark against the findings of Macomber and Yang (2012), who found that participants preferred higher sketch quality but lower render quality. Kendall's Tau-b was also used for this data set.

There were varied correlations between sketch/render quality levels and participants' ranking of cylinders based on sketches only ($\tau = 0.025$, $p = 0.254$), renders only ($\tau = -0.082$, $p < 0.001$), and combined quality ($\tau = -0.029$, $p = 0.065$). Reminder that ranking levels are from 1 (best) to 4 (worst), while the sketch and/or render quality levels are organized from I (worst) to IV (best), so the correlation is negated, to map a lower average ranking (better) to a higher (better) sketch and/or render quality level. The results for the cylinder benchmark condition indicate that participants preferred lower render quality based on the very weak statistically significant negative correlation, matching the Macomber and Yang study; but there was no statistically significant preference for any particular level of sketch quality.

4.1.2 Correlations between participant rankings and sketch and/or render quality levels using average ranking

It should be noted that Macomber and Yang analyzed their data using average rankings for each level of render and sketch, rather than correlation analysis, as was done in the present work. This prompted an additional analysis on the product groups (coffee, scissors, washer) using the "average ranking" method to determine if the lack of significant correlations was due to a non-linear or nonexistent relationship between the independent and dependent variables. A Mann–Whitney U test was also used to determine if the average rankings for any sketch and/or render quality levels were significantly different from one another (Clark-Carter 1997). If the p value < 0.005, the data from the two compared levels are different enough to validate results comparing those two levels against one another.

For the average ranking method, the data grouped all participant answers for the three product groups separated into function-based questions and preference-based questions. Figures 2, 3 show the mean rankings for each of these question types, comparing individual sketch quality and render quality levels to one another. The Mann–Whitney U analyses test these mean ranking, and these results are shown in Tables 3, 4, 5 and 6 with significant statistical difference in results are in gray. Only mean rankings with statistically different comparisons are discussed.

The results indicate several statistically significant differences among sketch and render quality levels. For function-based questions, participants preferred sketch and render quality level II as they were ranked lower than the other levels. For preference-based questions, participants preferred sketch and render quality levels II and III as they were lower ranked than levels I and IV. These results indicate that there were some non-linear trends in participant responses to varying sketch and render quality levels. Thus, though Hypothesis 1 was not supported, significant effects of sketch and render quality levels were observed. These results will be explored in the discussion section.

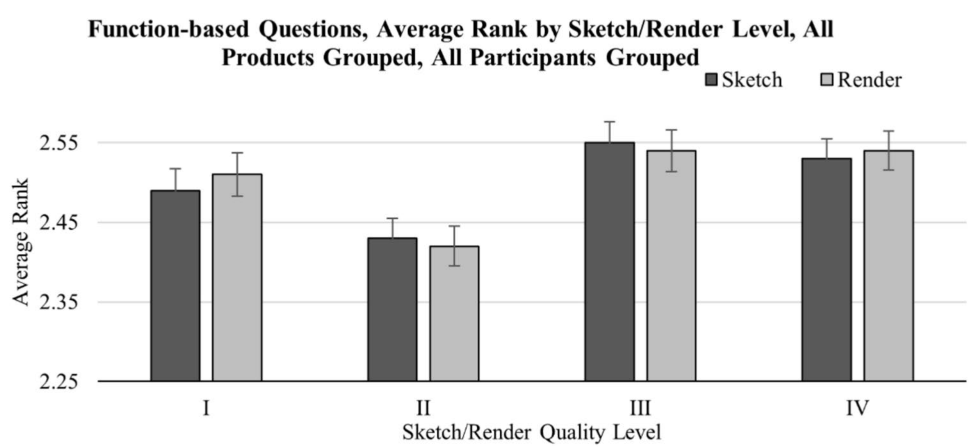


Fig. 2 Function-based questions, average rank by sketch and/or render quality level, all three products (coffee, scissors, washer) grouped, all participants grouped

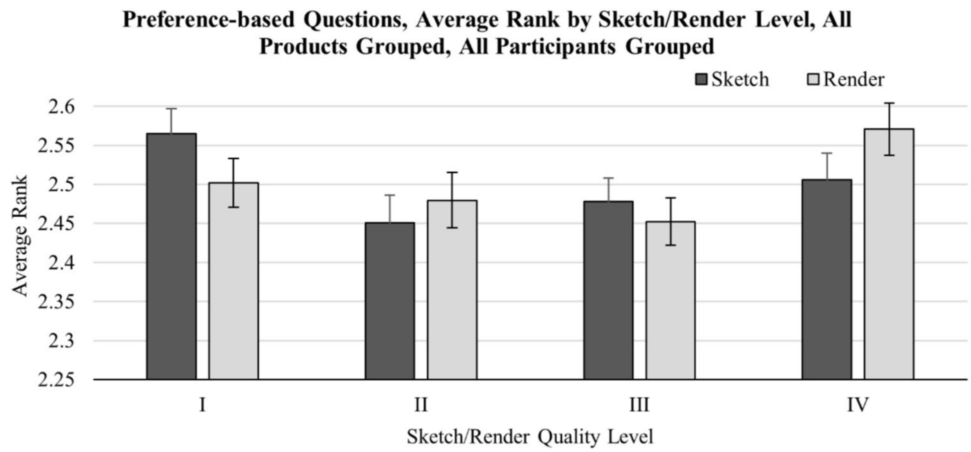


Fig. 3 Preference-based questions, average rank by sketch and/or render quality level, all three products (coffee, scissors, washer) grouped, all participants grouped

Table 3 Mann–Whitney *U* coefficient and *p* value for function-based questions from the coffee, scissors, and washers product groups, looking for significance between sketch quality levels

			Function-based questions (all products grouped)		
			Sketch quality level		
			II	III	IV
Sketch quality level	I	U	1,716,426.50	1,718,600.50	1,732,818.50
		p	0.122	0.155	0.299
	II	U		**1,661,837.50**	**1,670,473.50**
		p		**0.002**	**0.003**
	III	U			1,749,518.50
		p			0.665

Statistically significant comparisons are in bold

Table 4 Mann–Whitney *U* coefficient and *p* value for function-based questions from the coffee, scissors, and machines product groups, looking for significance between render quality levels

			Function-based questions (all products grouped)		
			Render quality level		
			II	III	IV
Render quality level	I	U	**1,728,234.00**	1,779,268.00	1,780,575.00
		p	**0.018**	0.415	0.456
	II	U		**1,698,491.00**	**1,692,713.50**
		p		**0.001**	**0.001**
	III	U			1,803,245.00
		p			0.934

Statistically significant comparisons are in bold

4.1.3 Correlations between participant rankings and baseline answers (user ratings) using Kendall's Tau-b

Table 5 Mann Whitney *U* coefficient and *p* value for preference-based questions from the coffee, scissors, and machines product groups, looking for significance between sketch quality levels

			Preference-based questions (all products grouped)		
			Sketch quality level		
			II	III	IV
Sketch quality level	I	U	**589,140.5**	**1,529,361.00**	665,931.50
		p	**0.018**	**0.046**	0.214
	II	U		671,480.00	608,102.50
		p		0.535	0.246
	III	U			738,159.50
		p			0.548

Statistically significant comparisons are in bold

Table 6 Mann Whitney *U* coefficient and *p* value for preference-based questions from the coffee, scissors, and machines product groups, looking for significance between render quality levels

			Preference-based questions (all products grouped)		
			Render quality level		
			II	III	IV
Render quality level	I	U	625,291.00	737,392.50	669,405.00
		p	0.625	0.254	0.118
	II	U		681,529.0	605,458.50
		p		0.593	0.064
	III	U			**712,334.50**
		p			**0.008**

Statistically significant comparisons are in bold

Next, all ranking data for function-based questions were compiled for all participants, and Kendall's Tau-b was used to determine if participant responses were significantly correlated with the baseline answers. Cylinders were not included because they did not have function-based questions. Results were separated by sketch quality levels only, render quality levels only, and sketch and render quality levels combined to see if sketch or render depictions impacted perception.

There was a statistically significant positive correlation between baseline answers and participants' perception of product functionality across all sketch quality ($\tau = 0.185$, $p < 0.001$), render quality ($\tau = 0.205$, $p < 0.001$), and combined quality levels ($\tau = 0.195$, $p < 0.001$). This indicates that participants were able to successfully judge product function characteristics as relative to the baseline answer.

4.2 Research question 2: differences between novice and advanced student designers

The second research question focused on testing if there were differences in the occurrence and intensity of illusory correlation for participants of different design expertise levels. The second hypothesis (H2) states that novice student designers are expected to develop more frequent and intense illusory correlations since they lack expertise to comprehensively evaluate a design concept. This would be indicated by stronger correlations between sketch and/or render quality and participant perceptions of functionality exhibited by novices compared to the advanced student designers. The participant data are split between novice and advanced participant responses and then compared to one another.

Since the sample size of novice and advanced student designers is different, Fisher Z analysis is used in both sections to confirm the validity of comparing two correlation coefficients from Kendall's Tau-b. It is important to note that the two correlation coefficients must be statistically significant for Fisher Z to be used. Therefore, Kendall's Tau-b coefficients that are not significant for both groups are not compared.

The first section looks for and compares the presence of illusory correlation between novice and advanced student designers in the product groups for sketch and/or render quality levels with both function-based and preference-based questions combined. The cylinder group is used as a benchmark group devoid of function to see if the product group results are similar. For the product group data set, if a novice designer has a significant stronger correlation than advanced student designers, then H2 is supported because novice student designers show stronger alignment between rankings and depiction levels.

The second section looks for and compares the presence of illusory correlation between the novice and the advanced student designers' correlations in perceived functionality. The results for the product groups function-based questions are combined and the compared to the baseline answers. The cylinder group was not included because it does not have function-based questions. If there is a significant positive stronger correlation for the advanced student designers than the novice student designers, then H2 is supported because advanced student designers are able to perceive functionality more consistently than novice student designers irrespective of sketch and/or render quality level.

4.2.1 Comparison of correlations between participant rankings and sketch and/or render quality levels for novice and advanced student designers

The data for all product groups and function and preference-based questions were separately combined for novice student designers and advanced student designers. The same was done for the cylinder groups preference-based questions, creating a novice and advanced student designers data set. Then the sketch quality levels only, render quality levels only, and both sketch and render quality levels combined were compared to one another for these four sets. Kendall's Tau-b was used to determine if higher rankings overall were associated with higher sketch and/or render quality levels of product depictions for advanced student designers compared to novices, and Fisher-Z was used to determine if a significant coefficient from novice student designers could be compared to a corresponding significant coefficient from advanced student designers.

The results of this test are shown in Fig. 4 following the negation of correlation to map a better ranking (1 is best) to a higher sketch/render quality level (IV). The results for all products grouped are not statistically significant, so they cannot be used toward supporting or refuting Hypothesis 2. For the cylinder data set, there are statistically significant weak negative correlations for render quality levels for both novice ($p = 0.015$) and advanced student designers ($p = 0.001$), which aligns with findings in RQ1 and of Macomber and Yang (2012) that participants preferred lower render quality. However, the Fisher Z analysis did not indicate that the correlation coefficients for novices and advanced student designers were statistically significantly different from one another ($z = 1.415$, $p = 0.157$), failing to support Hypothesis 2.

4.2.2 Comparison of correlations between participant rankings and baseline answers (user ratings) for novice and advanced student designers using Kendall's Tau-b

Next, all data were compiled, separated by level of expertise, for all products and all function-based questions in both the

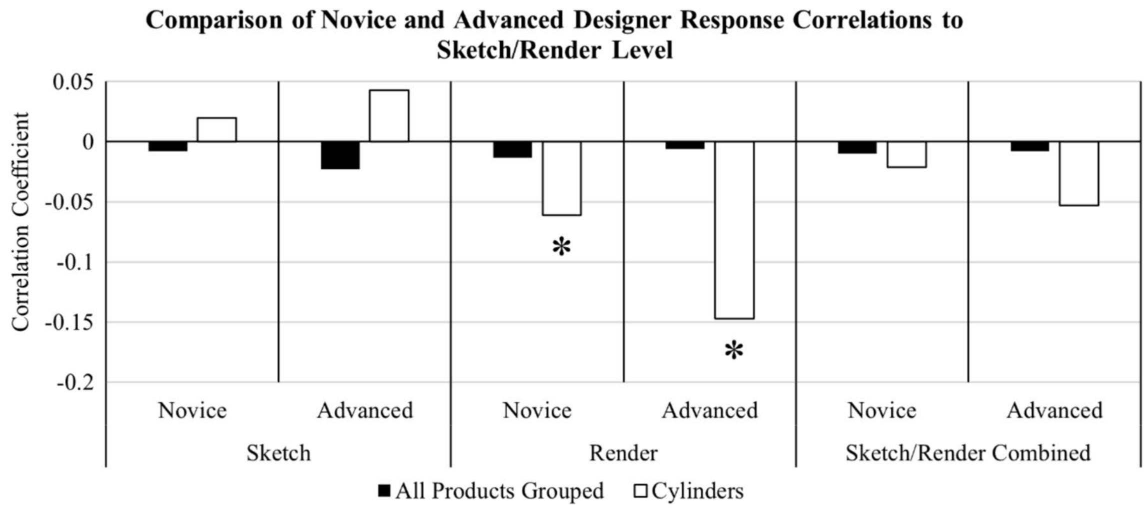


Fig. 4 Novice and advanced designer response correlations comparing all questions grouped as the three product groups to all the questions for the cylinders, looking at correlations between sketch quality levels only, render quality level only, and combined sketch and render quality levels; *indicates statistical significance at the $p \leq 0.05$ level

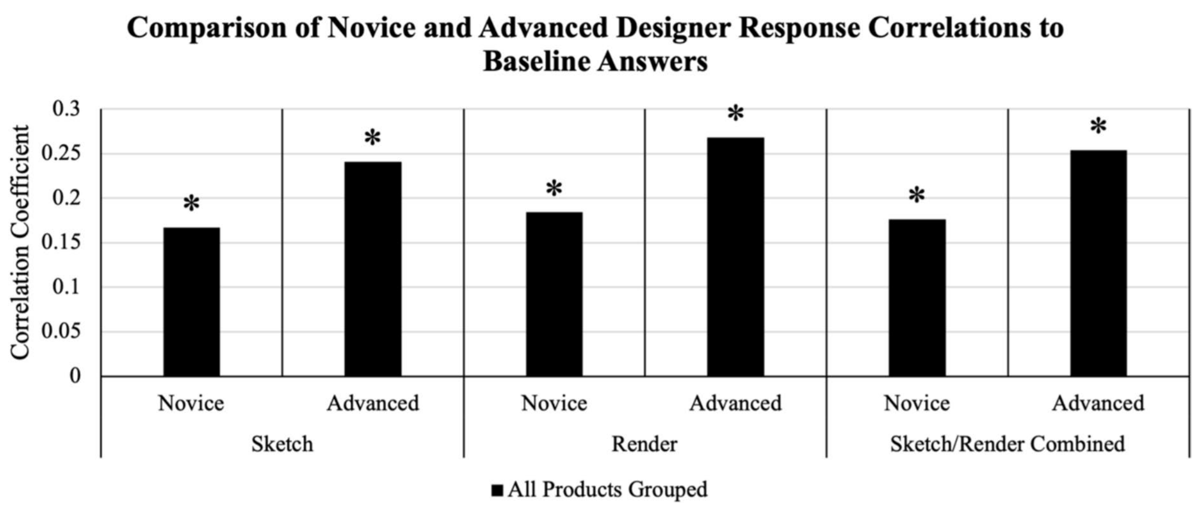


Fig. 5 Novice and advanced designer response correlations to baseline answers; *indicates statistical significance at the $p \leq 0.05$ level

sketch quality level and render quality level studies, comparing participant responses to the baseline answer. Cylinders are not included because they do not have function-based questions. Similarly to the section above, Kendall's Tau-b was then run on the data to determine if participant responses were significantly correlated with baseline answers, for novice and/or advanced student designers, with Fisher Z checking for significance of corresponding correlations.

The results of this test are shown in Fig. 5, looking for consistency or differences based on sketch and/or render use in depictions. For novices, the comparison between participant rankings and baseline answers for sketch quality level ($\tau = 0.167$, $p < 0.001$), render quality level ($\tau = 0.184$, $p < 0.001$), and combined levels ($\tau = 0.176$, $p < 0.001$) shows consistent statistically significant positive correlations. For advanced student designers, the comparison between participant rankings and baseline answers for sketch quality level ($\tau = 0.241$, $p < 0.001$), render quality level ($\tau = 0.268$, $p < 0.001$), and combined levels ($\tau = 0.254$, $p < 0.001$) also shows consistent positive correlations that are statistically significant. The Fisher Z analysis showed that these correlations were significantly stronger for advanced student designers compared to novice student designers for sketch quality level ($z = -2.868$, $p = 0.004$), render quality level ($z = -3.311$, $p = 0.001$), and combined sketch and/or render quality levels ($z = -4.312$, $p < 0.001$). The correlations within novice and advanced student designers' results for sketch quality levels, render quality levels, and sketch and render quality levels combined do not indicate noticeable differences based on sketch or render depictions. This finding supports Hypothesis 2 as the advanced student designers showed evidence that they had a better understanding of and ability to judge the functionality of the products they evaluated, making them less vulnerable to illusory correlation bias.

5 Discussion

5.1 General illusory correlation bias findings

The main findings failed to support Hypothesis 1; overall, participants did not judge product concepts to be of higher functionality if they were depicted using higher sketch or render quality levels, indicating that illusory correlation bias did not impact participants' view of product functionality. There were no statistically significant linear correlations between participant responses and sketch quality and/or render quality level, with only the rendered cylinders having a statistically significant negative correlation across all participants. In fact, participants were able to detect the relative functionality of the presented product concepts, despite the varying sketch and render quality levels, as evidenced by significant positive correlations to baseline answers. When the average rankings for each sketch and/or render quality level were compared, it uncovered a non-linear trend of participant preference in preference-based questions toward sketch and render quality levels II and III, with a preference against levels I and IV. For function-based questions, participants again showed strong preference toward sketch and/or render quality level II. This indicates a non-linear relationship between participant preferences and the sketch and/or render quality level.

Although some researchers have observed a U-shaped relationship between the value of a stimulus and its ability to be remembered, indicating that more extreme values are more notable (Fine and Minnery 2009; Madan and Spetch 2012), the "Goldilocks effect" suggests that people show a preference toward intermediate stimuli rather than extremes (Kidd et al. 2012; Keane and Tibbits 2018). This preference against extreme options can impact a design team setting, in which team members prefer intermediate ideas and may be biased against particularly good- or poor-quality sketches or ideas. In design practice, this can mean that anything considered visually extreme is overlooked. However, findings from the comparison of novices and advanced student designers indicate that design education can counteract this bias as discussed in the following section.

5.2 Illusory correlation bias and design expertise

While neither novice nor advanced student designers showed statistically significant correlations between their responses and the sketch and/or render quality level of the product depictions, advanced student designers had statistically significantly higher correlations to the baseline answer than novice student designers. This supports Hypothesis 2 as the advanced student designers showed evidence that they had a better understanding of and ability to judge the functionality of the products they evaluated. The higher correlation values from advanced student designers' responses support that more engineering design education helps reduce advanced student designers' vulnerability to illusory correlation bias. This trend is also supported by literature in psychology, in which graduate students in a study answered questions more accurately than undergraduate students (Starr and Katkin 1969). Current design practice supports findings from the present study as design engineers with more experience are given more responsibility to teach and lead those with less experience. This also supports the ability of curricula to properly teach engineering design students to avoid illusory correlation bias when analyzing functionality.

6 Limitations and future work

A limitation of this work is that findings are restricted to the context in which they were gathered. Conclusions can only be drawn about the specific product groups and products that were compared in this study. The participants involved in the study were students at a major southeast US institution, limiting factors, such as geographic location and potential cultural differences, which might arise in other locations. In addition, all participants were engineering students, leading to the sample population having a distribution of race and gender that does not represent the greater public in the USA or the world. As students, the participants are not yet as experienced or trained as practicing expert designers; as such, the findings cannot be extended to design experts or recommendations for their design practice. In addition, a great deal more novice student designers were able to be recruited for this study, compared to advanced student designers. In general, this trend occurs often in research with human subjects. The greater the level of expertise, the harder to access and recruit participants. However, with each individual ranking of each product able to be counted as a single data point in the correlation calculations, the sample sizes were quite large as shown in the results section. Nonetheless, more even sample sizes among the two participant groups would have been preferable.

A potential avenue for future study includes a closer look at curriculum-specific content to determine if educational background plays a role in mitigating illusory correlation bias since participants came from different backgrounds and were in the early stages of their engineering degree. Further longitudinal study in which the same participants repeated the survey upon completion of their engineering design training would permit analysis of how their specific education experience impacted their answers. Additional insights on participants' decision-making process could also be gained through qualitative responses in a post-survey, which should be included in similar experimental procedures in future.

The survey deployed to collect data was designed to randomize the order of appearance of sketch quality level and render quality level, as well as product groups and individual products, in an attempt to evenly distribute the effects of attention fatigue over the course of the study on the results gathered. Nonetheless, the surveys were long with the calculated time to complete it being 40.2 min with a standard deviation of 14.1 min. Data collections lasting this long with repetitive tasks can lead to fatigue and lowered attention to answers, which could potentially corrupt the fidelity of the data when compared to real-life decision-making scenarios.

The way the data were collected also differed from a real-life decision-making scenario based on the ways in which designers evaluate and select product concepts during the design process. Often, design teams select concepts in group sessions with active discussion and negotiation about their judgments of ideas. Making these decisions in a team setting is significantly different than making these decisions individually and then aggregating them. For example, in team-based discussions, other biases can occur, such as the bandwagon effect, in which team members are prone to agree with one another, regardless of their own individual opinion (Leibenstein 1950). Future work could examine the effect of group concept evaluation on the occurrence of illusory correlation bias as compared to individual concept evaluation.

The timing of participants' responses is another factor to consider. Based on the literature, intuitive thinking occurs as an immediate response, while analytical thinking takes over after a given amount of time (over 30 s) (Dane and Pratt 2009; Dane et al. 2012). Researchers cannot know empirically what type of thinking is occurring in the brains of the participants but rather can only assume based on the time taken to reach a decision. There were not enough data to analyze the impact of timing to compare intuitive and analytical thinking. Additionally, some participants did not complete all rankings for some individual questions, leaving one or more rankings blank and leading to incomplete data. This is the reason for the varying sample sizes among the different product groups, sketch and/or render quality levels, and expertise compared.

One can imagine that the variability in depiction of concepts in real-world settings would be much greater than that tested in this study simply due to multiple individuals generating the depictions. In this study, one designer drew all depictions used, leading to a level of uniformity in style. Ideally, this lends continuity to the study design as it removes a confounding factor from the experimental results. However, in real-world design settings, stylistic variances are bound to occur. Future work could examine the effect of stylistic differences on the occurrence of illusory correlation bias. In addition, analysis of the time taken to create the sketch and/or renders of the products could be analyzed to determine if there is a significant correlation with participant perceptions of functionality. This could have implications for how designer resources can be optimally utilized during visual communication stages of the design process.

Finally, there may have been an effect of the unusualness or familiarity with individual products within the product groups that could have impacted the occurrence or intensity of illusory correlation bias in a way that was not measured in this study. Future work could isolate this factor and study it specifically to determine if it has an impact.

7 Conclusions

This work explored how illusory correlation can impact a designer's perception of the functionality of a product during concept evaluation. In contrast with prior work, this study expanded beyond preference-based questions to explore participants' perceptions of the functionality of a product in order to evaluate how illusory correlation may impact decision-making in concept evaluation. It also features a novel comparison of designers' perceptions of functionality to actual ratings from customers who used the product. These contributions were achieved by addressing the following research questions:

RQ1. How does render and sketch quality level impact a designer's perception of the functionality of a design concept?

It was expected that illusory correlation would cause designers to perceive designs as more functional and exhibit a higher preference for designs that are presented with higher sketch and render quality. However, there was not a statistically significant correlation between participants' ranking of the products and the sketch and/or render quality level. This contradicts the findings of Macomber and Yang, who found that design stakeholders exhibit higher preference for higher-quality sketches/renderings (2012). However, the study by Macomber and Yang was preference-based, where participants were asked to rank the images based on how much they "liked" each image. In the present study, participants ranked images based on various aspects of their perceived functionality, simulating decisions that are more similar to those that would be made in design concept evaluation. For preference-based questions, a non-linear trend was found for “middle ground” quality level images that has potential for further research, but this finding does not impact the scope of this study for perception of functionality. This finding is promising due to the implication that designers performing concept evaluation are able to avoid biases that may otherwise lead them to select the concept that is the most visually well-represented.

RQ2. How does level of education affect the presence or intensity of illusory correlation for designers?

Advanced student designers were expected to develop weaker illusory correlations due to their increased design knowledge and training. However, neither novice nor advanced student designers exhibited a statistically significant correlation between product ranking and sketch and/or render quality level for either group. When comparing participants’ perceptions of functionality to a baseline answer established from actual user reviews, the advanced student designers’ responses were more strongly correlated to the baseline answer compared to the novices. This indicates that the additional design training and experience held by the advanced group resulted in a better intuitive understanding of the functionality of a product, reducing their vulnerability to biases such as illusory correlation bias. These results emphasize the importance of building design intuition in early design engineering coursework and training.

Appendix

Appendix 1. Sketch and render quality levels for all product groups

See Figs. 6, 7, 8, 9, 10, 11, 12.

Fig. 6 Coffee Maker Depictions—all sketch quality levels for each product

Fig. 7 Coffee Maker Depictions—all render quality levels for each product

Fig. 8 Washer Depictions—all sketch quality levels for each product

Fig. 9 Washer Depictions—all render quality levels for each product

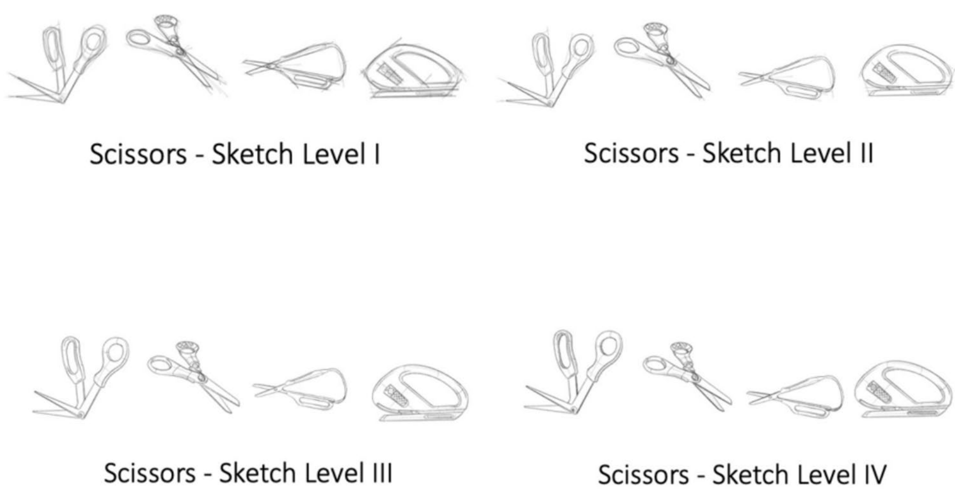


Fig. 10 Scissors Depictions—all sketch quality levels for each product

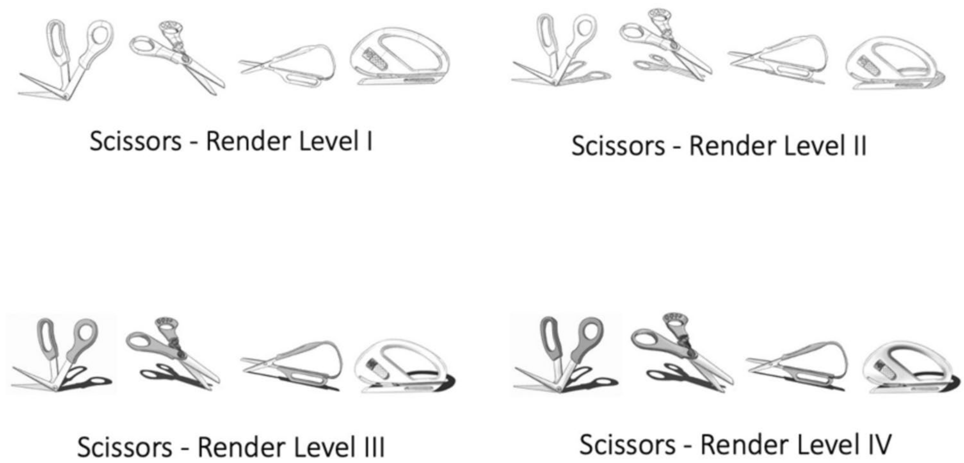


Fig. 11 Scissors Depictions—all render quality levels for each product

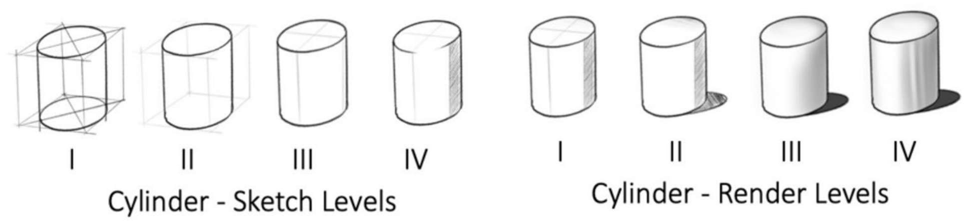


Fig. 12 Cylinder depictions—all sketch and render quality levels

Appendix 2. Product attribute questions used within the survey

See Table 7 and Fig. 13.

Table 7 Product Attribute Questions used within the Survey; 6 generic functionality, 2 product group specific functionality, and 2 preference questions total for each of the three product groups; no functionality and 4 preference questions for the cylinder group

Functionality (6): Generic for Coffee, Washer, Scissors	For all product groups: Which product is the most comfortable for the user for a singular use? Which product is the most intuitive to use? Which product has the easiest maintenance for the user? Which product uses the least amount of input resources for the most amount of output? (i.e., highest efficiency) Which product is most expensive to buy for the consumer? Which product functions most effectively?
Functionality (2): Product Group Specific for Coffee	Which product makes the cup with the highest caffeine content? Which product allows the user to make a cup of coffee the fastest?
Functionality (2): Product Group Specific for Washer	Which product is best for gentle washing of clothes? Which product is best for cleaning a larger variety of types of clothing?
Functionality (2): Product Group Specific for Scissors	Which product is best for use to cut multiple pieces of paper at one time? Which product is best for use to cut complex shapes with precision?
Preference (2): Generic for Coffee, Washer, Scissors	Assuming the products are priced as they would be for purchase in a store, which product would you purchase? Assuming all prices are equal for purchase, which product would you purchase?
Preference (4): Cylinders	Which drawing is the most professional? Which drawing took the longest time to sketch? Which drawing is the most clear? Which drawing do you like best?

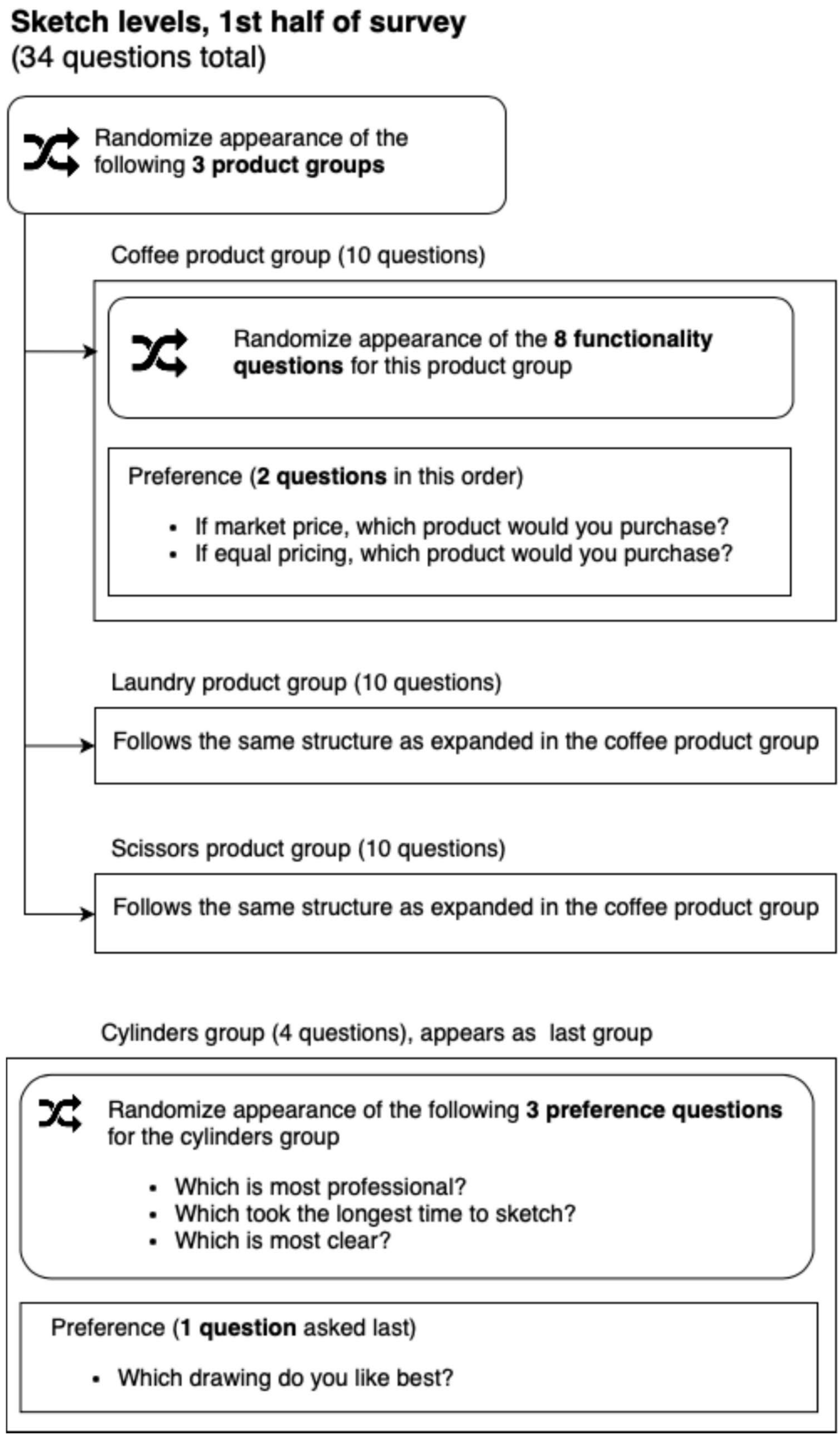


Fig. 13 Order of the product attribute questions in the first half of the survey as the example; the first half contains sketch quality level product depictions only and the second half is the render quality level product depictions following the same shown in this visual

Acknowledgements The authors would like to thank Dr. Kenton Blane Fillingim and Dr. Ricardo Bonilla-Alicea for their input and guidance in the creation of the study, and Varun Nambiar for his assistance in the development of the product depictions for this study.

Author contributions N.D. and A.S wrote the main manuscript text and H.S. prepared the literature review. All authors reviewed the manuscript.

Funding National Science Foundation, CMMI 1846048, CMMI 1846048, CMMI 1846048.

Data availability Data collected and analyzed in this study is available from the corresponding author upon reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Baylor AL (1997) A three-component conception of intuition: Immediacy, sensing relationships, and reason. New Ideas Psychol 15:185–194. https://doi.org/10.1016/S0732-118X(97)00016-0

Blanco F (2017) Cognitive bias. In: Vonk J, Shackelford T (eds) encyclopedia of animal cognition and behavior. Springer International Publishing, Cham, pp 1–7

Botsch RE (2011) Chapter 12. Significance and measures of association

Cham JG, Yang MC (2008) Does Sketching Skill Relate to Good Design? American Society of Mechanical Engineers Digital Collection, pp 301–308

Chapman LJ (1967) Illusory correlation in observational report. J Verbal Learn Verbal Behav 6:151–155. https://doi.org/10.1016/S0022-5371(67)80066-5

Chapman LJ, Chapman JP (1967) Genesis of popular but erroneous psychodiagnostic observations. J Abnorm Psychol 72:193–204. https://doi.org/10.1037/h0024670

Clark-Carter D (1997) Doing quantitative psychological research: From design to report. Psychology Press/Erlbaum (UK) Taylor & Francis, Hove, England

Crismond D (2001) Learning and using science ideas when doing investigate-and-redesign tasks: a study of naive, novice, and expert designers doing constrained and scaffolded design work. J Res Sci Teach 38:791–820. https://doi.org/10.1002/tea.1032

Cross N (2001) Chapter 5 - design cognition: results from protocol and other empirical studies of design activity. In: Eastman CM, McCracken WM, Newstetter WC (eds) Design knowing and learning: cognition in design education. Elsevier Science, Oxford, pp 79–103

Cross N (2004) Expertise in design: an overview. Des Stud 25:427–441. https://doi.org/10.1016/j.destud.2004.06.002

Cross N, Christiaans H, Dorst K (1994) Design expertise amongst student designers. J Art Design Education 13:39–56. https://doi.org/10.1111/j.1476-8070.1994.tb00356.x

Dane E, Pratt MG (2009) Conceptualizing and measuring intuition: a review of recent trends. International review of industrial and organizational psychology. John Wiley & Sons Ltd, pp 1–40

Dane E, Rockmann KW, Pratt MG (2012) When should I trust my gut? Linking domain expertise to intuitive decision-making effectiveness. Organ Behav Hum Decis Process 119:187–194. https://doi.org/10.1016/j.obhdp.2012.07.009

Deininger M, Daly SR, Lee JC et al (2019) Prototyping for context: explorig stakeholder feedback based on prototype type, stakeholder group and question type. Res Eng Design 30:453–471. https://doi.org/10.1007/s00163-019-00317-5

El Dehaibi N, Goodman ND, MacDonald EF (2019) Extracting customer perceptions of product sustainability from online reviews. ASME J Mech des 141:12. https://doi.org/10.1115/1.4044522

Fiedler K, Russer S, Gramm K (1993) Illusory correlations and memory performance. J Exp Soc Psychol 29:111–136. https://doi.org/10.1006/jesp.1993.1006

Fillingim KB, Shapiro H, Fu KK (2023) Error management bias in student design teams. ASME J Mech des 145:10. https://doi.org/10.1115/1.4055899

Fine MS, Minnery BS (2009) Visual salience affects performance in a working memory task. J Neurosci 29:8016–8021. https://doi.org/10.1523/JNEUROSCI.5503-08.2009

Gilovich T, Griffin D, Kahneman D (eds) (2002) Heuristics and biases: the psychology of intuitive judgment. Cambridge University Press, Cambridge

Häggman A, Tsai G, Elsen C et al (2015) Connections between the design tool, design attributes, and user preferences in early stage design. J Mech Design. DOI 10(1115/1):4030181

Hallihan GM, Shu LH (2013) Considering Confirmation bias in design and design research. J Integr des Process Sci 17:19–35. https://doi.org/10.3233/jid-2013-0019

Hallihan GM, Cheong H, Shu LH (2013) Confirmation and cognitive bias in design cognition. American Society of Mechanical Engineers Digital Collection, pp 913–924

Hamilton DL, Gifford RK (1976) Illusory correlation in interpersonal perception: a cognitive basis of stereotypic judgments. J Exp Soc Psychol 12:392–407. https://doi.org/10.1016/S0022-1031(76)80006-6

Hamilton DL, Dugan PM, Trolier TK (1985) The formation of stereotypic beliefs: further evidence for distinctiveness-based illusory correlations. J Pers Soc Psychol 48:5–17. https://doi.org/10.1037/0022-3514.48.1.5

Han J, Forbes H, Schaefer D (2021) An exploration of how creativity, functionality, and aesthetics are related in design. Res Eng Design 32:289–307. https://doi.org/10.1007/s00163-021-00366-9

Hartsough WR (1975) Illusory correlation and mediated association: a finding. Can J Behav Sci Revue Canadienne des Sci Du Comportement 7:151–154. https://doi.org/10.1037/h0081903

Jenkins HM, Ward WC (1965) Judgment of contingency between responses and outcomes. Psychol Monogr Gen Appl 79:1–17. https://doi.org/10.1037/h0093874

Jiang H, Kwong CK, Yung KL (2017) Predicting Future importance of product features based on online customer reviews. J Mech des 139:10. https://doi.org/10.1115/1.4037348

Keane RG, Tibbits BF (2018) The goldilocks effect in naval ship design: too little, too much, just right. OnePetro

Kidd C, Piantadosi ST, Aslin RN (2012) The goldilocks effect: human infants allocate attention to visual sequences that are neither too

simple nor too complex. PLoS ONE 7:e36399. https://doi.org/10.1371/journal.pone.0036399
Koh ECY, De Lessio MP (2018) Fixation and distraction in creative design: the repercussions of reviewing patent documents to avoid infringement. Res Eng Design 29:351–366. https://doi.org/10.1007/s00163-018-0290-y
Koriat A, Lichtenstein S, Fischhoff B (1980) Reasons for confidence. J Exp Psychol Human Learn Memory 6:107–118. https://doi.org/10.1037/0278-7393.6.2.107
Leibenstein H (1950) Bandwagon, snob, and veblen effects in the theory of consumers' demand. Q J Econ 64:183–207. https://doi.org/10.2307/1882692
Li S, Liu F, Lu H et al (2021) Product family lean improvement based on matching deep mining of customer group preference. Res Eng des 32:469–488. https://doi.org/10.1007/s00163-021-00367-8
Linder A, Svensson MY (2019) Road safety: the average male as a norm in vehicle occupant crash safety assessment. Interdisc Sci Rev 44:140–153. https://doi.org/10.1080/03080188.2019.1603870
Macomber B, Yang M (2012) The role of sketch finish and style in user responses to early stage design concepts. American Society of Mechanical Engineers Digital Collection, pp 567–576
Madan CR, Spetch ML (2012) Is the enhancement of memory due to reward driven by value or salience? Acta Physiol (Oxf) 139:343–349. https://doi.org/10.1016/j.actpsy.2011.12.010
Munro GD, Stansbury JA (2009) The dark side of self-affirmation: confirmation bias and illusory correlation in response to threatening information. Pers Soc Psychol Bull 35:1143–1153. https://doi.org/10.1177/0146167209337163
Murphy RA, Schmeer S, Vallée-Tourangeau F et al (2011) Making the illusory correlation effect appear and then disappear: the effects of increased learning. Q J Exp Psychol 64:24–40. https://doi.org/10.1080/17470218.2010.493615
Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. Rev Gen Psychol 2:175–220. https://doi.org/10.1037/1089-2680.2.2.175
Starr BJ, Katkin ES (1969) The clinician as an aberrant actuary: illusory correlation and the incomplete sentences blank. J Abnorm Psychol 74:670–675. https://doi.org/10.1037/h0028466
Strigini L (1996) Engineering judgement in reliability and safety and its limits: what can we learn from research in psychology?
Tversky A, Kahneman D (1973) Availability: a heuristic for judging frequency and probability. Cogn Psychol 5:207–232. https://doi.org/10.1016/0010-0285(73)90033-9
Tversky A, Kahneman D (1974) Judgment under uncertainty: heuristics and biases. Science 185:1124–1131. https://doi.org/10.1126/science.185.4157.1124
Viswanathan VK, Linsey JS (2013) Role of sunk cost in engineering idea generation: an experimental investigation. J Mech des 135:121002
Ward WC, Jenkins HM (1965) The display of information and the judgment of contingency. Can J Psychol Revue Canadienne De Psychologie 19:231–241. https://doi.org/10.1037/h0082908