Paper to be presented at
the DRUID16 20th Anniversary Conference
Copenhagen, June 13-15, 2016

# Design myopia and vicarious learning from good versus bad examples: Evidence from creative design competitions

**Christoph Riedl**
**Northeastern University**
**D'Amore-McKim School of Business**
**c.riedl@neu.edu**

**Victor Seidel**
**Babson College**
**FW Olin Graduate School of Business**
**vseidel@babson.edu**

**Abstract**

High-quality creative designs create tremendous value for organizations, but how do individual designers learn to produce better designs? Learning often presumes evaluation of performance that is objective and immediate, but in creative design evaluation is social and temporally displaced, providing hard-to-interpret signals for learning. Drawing on data from a ten-year panel of almost 180,000 T-shirt design submissions and 150 million design evaluations on an online crowdsourcing platform, we investigate how individuals learn from their own work and vicariously learn from observing others' work. We find that in the absence of vicarious learning, individuals experience "design myopia" resulting in successively lower quality designs before reaching a positive learning rate. Furthermore, individuals learn from evaluating good examples of others, but they generally fail to learn from evaluating bad examples. We also find that experience helps individuals not only to gain high evaluations from others but also to learn to understand the "black box" of how designs are chosen for production. We discuss implications for the development of online crowdsourcing platforms and for the management of creative design more broadly.

Jelcodes:O31,-

# Design myopia and vicarious learning from good versus bad examples: Evidence from creative design competitions


## ABSTRACT

High-quality creative designs create tremendous value for organizations, but how do individual designers learn to produce better designs? Learning often presumes evaluation of performance that is objective and immediate, but in creative design evaluation is social and temporally displaced, providing hard-to-interpret signals for learning. Drawing on data from a ten-year panel of almost 180,000 T-shirt design submissions and 150 million design evaluations on an online crowdsourcing platform, we investigate how individuals learn from their own work and vicariously learn from observing others' work. We find that in the absence of vicarious learning, individuals experience "design myopia" resulting in successively lower quality designs before reaching a positive learning rate. Furthermore, individuals learn from evaluating good examples of others, but they generally fail to learn from evaluating bad examples. We also find that experience helps individuals not only to gain high evaluations from others but also to learn to understand the "black box" of how designs are chosen for production. We discuss implications for the development of online crowdsourcing platforms and for the management of creative design more broadly.

## INTRODUCTION

We present a study of how individuals involved in creative design competitions hosted on an online crowdsourcing platform learn from a combination of direct and indirect experience. High-quality creative design work can create tremendous value for organizations. High quality designs help technical products gain acceptance (Hargadon & Douglas, 2001), are the main basis for competition in cultural markets (Wijnberg & Gemser, 2000), and provide individuals with aesthetic value (Eisenman, 2013; Rindova & Petkova, 2007). There has been mounting interest in the use of designers by organizations as a source of value creation (Ravasi & Lojacono, 2005; Rindova, Dalpiaz, & Ravasi, 2011; Rindova & Petkova, 2007). Despite the immense value that designers may bring to organizations, it is unclear how designers learn to produce better designs. Furthermore, it is unclear how information systems intended to facilitate learning and improve retention of contributors should be designed.

Prior studies have pointed to experience as central for individual learning, whether it is direct experience (learning by doing) or indirect experience (vicarious learning by observing the work of others; (e.g. KC, Staats, & Gino, 2013; Newell & Rosenbloom, 1981). However, learning through either direct or indirect experience often presumes an evaluation of one's performance that is objective and immediate. Yet, in design work, evaluation is social and often temporally displaced (Bloch, 1995; Eisenman, 2013), providing potentially unclear signals for learning. Indeed, in some innovative work, such as in suggesting product ideas on a online crowdsourcing platform, there is no evidence of learning from experience (Huang, Singh, & Srinivasan, 2014). As a result, the degree to which individual-level learning can be achieved in design work—a process that depends so critically on individual creativity—requires investigation.

In this study we quantify the rate of learning of individuals engaged in creative design competitions. We analyze the direct and indirect learning activities of over 55,000 individuals making almost 180,000 T-shirt design submissions to an online crowdsourcing platform, hosting creative design contests over a ten year period. Many prior studies of creative design work have been focused on in-depth qualitative studies (e.g. Hargadon & Sutton, 1997; Harrison & Rouse, 2015), while in contrast our dataset allows us to very precisely quantify learning across a large sample. Our dataset contains almost 150 million evaluations that individuals made on others' work, offering a detailed window into the role of social learning by observing and evaluating others' designs.

Our empirical analyses allowed us to quantify the learning of individuals due to both direct and indirect experience. We have five main findings. First, we find that individuals doing creative design work learn from both direct and indirect experience. Second, we find that in the absence of vicarious learning, individuals experience a period of *decreased* performance prior to positive returns. In our empirical setting we find it would take about seven design submissions before positive learning effects accrue, which we label as "design myopia." Third, we find strong complementarities between direct and indirect experience that can help individuals overcome design myopia. Fourth, we find that in creative design work, individuals learn from good examples but fail to learn from bad examples. However, individuals can overcome their inability to learn from bad examples through prior experience in evaluating good examples. Fifth, we find that learning from direct and indirect experience not only helps individuals produce designs of high market potential but also helps them learn about the "black box" of how firms choose winners of design competitions.

Our findings advance theories of learning in the context of creative design work, and we make three primary contributions. First, by drawing on large-scale empirical data, we are able to precisely quantify learning curves that cover both direct and indirect experience, and in doing so specify unique mechanisms in creative design work—such as design myopia and vicarious learning from good examples—within this setting. Secondly, we contribute a more nuanced view of the role of failure in creative work. Most of the work on learning from failure in design settings has been based on failures observed from one's own direct experience.  Thirdly, our study provides a means to understand how creative design work relates to, yet differs from, other forms of work in the shape and rate of the learning curve, providing a basis for further research in this area in both offline and online contexts.

## LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Examples abound of how creative design work can create significant value. Thomas Edison was not only a great inventor of new technologies and functions but also importantly used design to give form to electric light fixtures that mimicked the established gas lighting system to facilitate adoption (Hargadon & Douglas, 2001). Designers at Apple Inc. changed the perception of home computers by replacing dreary beige boxes with colorful translucent housings (Ravasi & Lojacono, 2005), and designers at the Italian firm Alessi advanced their market share of household goods through design innovations (Rindova et al., 2011).

### The nature of creative design work

Herbert Simon noted that as opposed to focusing on how things are, “Design … is concerned with how things ought to be, with devising artifacts to attain goals” (Simon, 1996, p. 114). The outcome of design work is a new artifact, and the quality of this artifact is measured by the social evaluation of others rather than by some pre-existing objective criteria. Capturing this nature of judging and evaluating a design, Stefano Giovannoni, who has designed many objects for Alessi, described design in the following context: “Objects are not beautiful or ugly but are either suited or not to their time” (Fiell & Fiell, 2005, p. 95). While individual designers need to learn specific skills to produce a design, it is critical that they also learn to understand shifting consumer preferences and how to suit a design to a specific moment in time. In summary, design performance is not technically measured to a specification but is a social evaluation made by consumers.

Designers have more control over their efforts than others involved in the innovation process such as mechanical or electrical engineers. While the components of an integrated circuit may be the result of many different efforts, a new product form often comes from the work of a single designer. Indicative of this individual focus, a review of design at the start of the 21st century focused on the work of individual designers (Fiell & Fiell, 2005), including Ross Lovegrove working with Japan Airlines, J. Mays working with Ford Motor Company, and Ron Arad working with Vitra. While designers often have to integrate their work with that of engineering, marketing managers, and others, they frequently maintain their individual stamp on the final design. Increasingly, creative design work can be sourced through online crowdsourcing platforms (e.g. Seidel & Langner, 2015) that provide a new way to involve individuals in creative design work. We next turn to the focus of our study on how individuals learn to improve their design performance on these platforms.

**The role of direct experience**

The role of direct experience at the individual-level has been demonstrated in a variety of settings, often portrayed as a learning curve (also called an experience curve) of increasing performance over time, though with diminishing returns. Individual learning curves under direct experience have been charted in settings ranging from cigar manufacturing to object identification tasks (For one influential review, see Newell & Rosenbloom, 1981). While the role of direct experience in learning manufacturing tasks is well-established (Argote & Epple, 1990), and is also seen in knowledge work (Boh, Slaughter, & Espinosa, 2007), many current accounts use organizational data rather than examining fundamental individual effects, though individual learning is the basis for organizational learning (Argote & Miron-Spektor, 2011). Studies demonstrating the effect of direct experience on learning feature a variety of technical measures of performance that are used to evaluate learning, such as labor cost to produce a part (Wright, 1936), surgical operation procedure time (Pisano, Bohmer, & Edmondson, 2001), or game score (Schilling, Vidal, Ployhart, & Marangoni, 2003).

It is unclear whether direct experience plays the same role in creative design work. Learning in creative design work requires gaining understanding of market potential and adapting to changing consumer preferences rather than merely meeting a fixed technical measure of performance. For example, the work of a manufacturing operator can be evaluated on how she improves the time to complete a task on a machine to a fixed level of quality (Newell & Rosenbloom, 1981). In contrast, consumer preferences are dynamic, shifting often quickly under the effect of changing cultural tastes, societal values, consumer experience, and fashion trends (Bloch, 1995).

Given that the measure of what constitutes a good design is not fixed but is continuously socially evaluated, it could be that experience from producing one design has little impact on one's ability to improve the quality of the next design. Indeed, in some cultural markets, past direct experience has been found to be a poor predictor of future success (Seifert, Siemsen, Hadida, & Eisingerich, 2015). If this is the case, rather than direct experience leading to learning, past direct experience could be viewed as misleading in the search for truly novel designs and innovations. Other studies also support skepticism that there can be learning in creative work. A study that drew on the Dell IdeaStorm.com online platform for sourcing new product ideas looked for learning curve effects, but this work found that there was no evidence of learning from experience (Huang et al., 2014). Another study concluded that in innovative work, increasing experience leads to more incremental outcomes (e.g. Audia & Goncalo, 2007), which implies a firm might benefit from seeking newcomers for innovation instead of trying to develop the performance of its own designers.

Counter to these studies on the limits of learning, there is evidence that direct experience could lead to learning even in creative work. For example, in a qualitative study of creative design teams, the direct experience of producing a design and obtaining direct feedback has been found to provide opportunity to learn (Harrison & Rouse, 2015), and a lab study in the realm of creative design work investigated how direct experience in designing new origami objects would improve design quality (Gino, Argote, Miron-Spektor, & Todorova, 2010). On balance, and despite some evidence calling direct experience into question in creative efforts, we hypothesize that there will be learning effects due to direct experience, which we can state as follows:

*Hypothesis 1: An individual's prior direct experience has a positive effect on an individual's current design performance.*

**Vicarious learning through indirect experience**

Learning is advanced not only by incorporating feedback from direct experience but also from the vicarious learning acquired through the indirect experience of observing others. As summarized in

Bandura's treatise on social learning theory, "…virtually all learning phenomena resulting from direct experience occur on a vicarious basis by observing other people's behavior and its consequences for them" (Bandura, 1977, p. 12). The role of vicarious learning is well established in studies of work in non-creative settings. For example, a study of surgeons completing procedures using minimally invasive cardiac surgery found that they learned from their indirect experience (KC et al., 2013). In addition to studies of individuals completing tasks, the role of vicarious learning has also been examined at the team level (e.g. Bresman, 2010) and organizational level (e.g. Boh et al., 2007; Pisano et al., 2001; Srinivasan, Haunschild, & Grewal, 2007; Wiersma, 2007).

Design work would also be expected to be shaped by vicarious learning through indirect experience. For example, a qualitative study within the product design firm IDEO found designers gained new insight in part by observing others performing design tasks (Hargadon & Sutton, 1997). Design education relies on the use of "critiques" where designers see each other's work being publically evaluated by their instructors and peers (Uluoğlu, 2000). Studies within social learning theory have noted that "observers generally learn faster than do the performers themselves" and that "this is especially true if the tasks depend more heavily on conceptual than on manual skills" (Bandura, 1977 , p. 122). The relative magnitude of learning through indirect experience when compared with direct experience is open to investigation; experimental studies of student teams doing creative work demonstrated a positive role for vicarious learning but one which was lower than gained by direct experience (Gino et al., 2010). Based on this prior work, we propose the following hypothesis for the role of indirect experience in creative design work:

*H2a: An individual's prior indirect experience in evaluating others' designs has a positive effect on an individual's current design performance.*

Prior studies of production work have shown that indirect and direct learning are complementary. For example, work of teams completing the tasks to license new pharmaceuticals demonstrated that vicarious learning activities are more strongly associated with performance when coupled with increased direct experience (Bresman, 2010). Similar effects have been shown in individual surgeons increasing their performance by combining insights from direct and indirect experience (KC et al., 2013).

Design work is defined by the synthesis of information, where designers seek to find "fit" of physical properties to revealed preferences among their target audience (Alexander, 1964), and so design work may be expected to see this reinforcement as well. As with other forms of innovation, design work relies on being able to creatively and continually synthesize ideas, concepts, and external inputs (Harvey, 2014). The focus on synthesizing information obtained directly and that obtained through observing others leads us to the hypothesis:

*Hypothesis 2b: An individual's past direct and indirect experience have a complementary effect on an individual's current design performance.*

**Observing and evaluating good and bad designs**

Vicarious learning comes from observing and evaluating the work of others, but these effects may depending on the quality of what is indirectly experienced; there may be different contributions to observing others' "good" versus "bad" creative design work. While most studies focused on the role of success or failure are rooted in learning from one's own direct experience (e.g. Sitkin, 1992), some studies have specifically investigated the role of others' good and bad examples on one's own learning. One study looked at how surgeons learned from the work of others, finding that surgeons learned more from others' failures than from others' successes (KC et al., 2013). One reason why individuals may learn more from bad examples rather than good examples relates to feelings of "psychological safety." Thus failures of others can be more easily absorbed since the climate in which the learning occurs feels safer (Edmondson, 1999).

Observing and evaluating others' bad designs could provide opportunities to learn what mistakes to avoid. Promoting "learning from failure" is a common mantra in creative design work (Hargadon & Sutton, 1997; Harrison & Rouse, 2015) as well as more generally (Firestein, 2015), and the opportunity to observe the failed designs of others could provide an efficient social learning approach to learn what not to do (Bandura, 1977). However, learning from failure may be more difficult in the context of design where the solution space is typically very large. There might simply be too many ways in which designs could fail, making it difficult to discern meaningful information from failed designs.

Observing and evaluating the good work of others can also have benefits. In a general sense, creative designers tend to appreciate others' good work. "Star" designers are featured in industry texts as ways to inspire other design work (e.g. Fiell & Fiell, 2005), and successful design firms are often those that gain attention through winning design awards for their efforts (Hargadon & Sutton, 1997). In a more commercial sense, by observing and evaluating the work of others, an individual has the opportunity to vicariously examine how others have interpreted current consumer preferences. Creative design requires not only skills of how to develop a new product form or graphic design but also an understanding of consumer preferences, which is a noisy signal consisting of a myriad of elements (Bloch, 1995). For example, one major analysis of how creative design work impacts purchasing intentions reviewed eight factors that can alter how a consumer responds to a new product design form, such as a consumer's cultural and social context, design acumen, and innate design preferences (Bloch, 1995). By observing a wide variety of others' good work, an individual is able to efficiently draw upon the collective experience of what others have learned about a market and how they may have integrated this into their own designs. Thus, good designs may in fact contain more information (i.e., have higher entropy) than bad designs (Firestein, 2015). Learning from good designs is thus likely more valuable than learning from bad designs because such good designs have already incorporated the noisy signals inherent in understanding consumer preferences (or at least a larger proportion of that noisy signal has been incorporated). This leads us to formally state a hypothesis as:

*Hypothesis 3a: An individual's prior indirect experience in evaluating others' high-performing designs has a greater effect on an individual's current design performance than does indirect experience in evaluating others' low-performing designs.*

A critical complication in the context of design is that at the time of observing and evaluating others' work, it may not be apparent how others are making their own evaluation. An individual's own appreciation for what is a good versus bad design is best informed by being exposed to a wide variety of good designs. More indirect experience gained from observing good examples could help individuals to more clearly evaluate which are bad examples and thus learn from them. Past indirect experience from exposure to good work may provide individuals with information necessary to help them better interpret bad work, putting bad work into context and then being able to better infer a lesson from the mistakes of others. Stated as a hypothesis:

*Hypothesis 3b: An individual's prior indirect experience in evaluating others' high- and low-performing designs will have a complementary (i.e., positive interaction) effect on an individual's current design performance.*

## METHODS AND MEASURES

### Research Setting

Data for our study comes from a ten year panel of t-shirt designs submitted to the Threadless website (www.threadless.com), a platform for t-shirt design competitions and e-commerce site. At the time of our study, Threadless hosted an online community of several thousand designers who could submit graphical designs for t-shirts. Our data contains all design submissions made by over 55,000 individuals during a ten year period from 2001 through 2011, consisting of almost 180,000 total submissions and 150 million evaluations of these designs. Since the notion of learning is only meaningful for individuals who make at least two design submissions, our panel regression includes 33,813 individuals with two or more design submissions. Obvious spam, blank, or duplicate submissions were filtered out by Threadless and are not put up for voting on the website and are not part of our dataset. Our dataset averages about 340 design submissions per week.

Every week, submitted designs were posted on the website for rating, and ratings came from community members (fellow designers as well as interested consumers) to evaluate the designs to be printed by the firm, using a scale from 0 to 5. Ratings were anonymous and no average score was shown at the time of voting—only a counter of how many ratings had been submitted for a design. This community rating serves as a sign of market potential (Huang et al., 2014) and is our dependent variable to assess learning.

Among a pool of the t-shirts that received the highest ratings, Threadless would typically pick three to six designs per week to be printed and subsequently sold through the Threadless e-commerce site. Decisions of which T-shirts were printed rested with the management team at Threadless. The management team would consider not only community evaluation of designs but also how to balance a portfolio of T-shirt themes, and other strategic factors—thus effectively curating a "house style" idiosyncratic to Threadless. Winning designers would receive a cash prize and store credit if their design was selected for printing. In effect, the community ratings represented a judgment-based evaluation of performance, based on a continuous variable of score from 0 to 5. The firm further used a choice-based evaluation of performance, choosing specific designs that were not necessarily the absolute top-rated score. We will show analyses that demonstrate separate mechanisms of learning and that learning accrues at both the community-level of judgment-based evaluation and the firm-level of choice-based evaluation.

**Dependent and independent variables**

*Dependent Variable:* $Score_{it}$. Our primary dependent variable $Score_{it}$ measures the performance of a design submitted by individual *i* at time *t*. Design performance is assessed on a scale from 0-5 with higher scores representing better designs. $Score_{it}$ is the mean of all community ratings a design received.

*Direct and indirect experience.* We hypothesized that prior design experience from directly engaging in design activity oneself or from indirectly observing others would affect current design performance. We measure prior direct design experience as the cumulative number of designs submitted by individual *i* prior to submitting the current design ($Exp_{it-1}$). This prior direct experience contains any design submission made, independent of how well or poorly it fared in a competition. Indirect experience is measured by the cumulative number of all prior evaluations cast by an individual $i$ before submitting the current design ($Votes_{it-1}$). The number of ratings an individual has cast serves as a proxy variable for the knowledge acquired vicariously through the indirect experience of observing and evaluating others' designs. To test the differential effect of voting on either a design of high performance or a design of low performance we create separate measures that count the respective numbers of ratings cast for high and low performance designs ($Votes_{it-1\,(Good)}$ and $Votes_{it-1\,(Bad)}$). The two separate counts added together yield the total indirect experience ($Votes_{it-1(Good)} + Votes_{it-1(Bad)} = Votes_{it-1}$). A design was considered to be of high performance if—ex-post considering all ratings that a design received—the average score of that design was higher than the mean score of all designs (i.e., *Score* greater than 1.91) otherwise it was considered of low performance.[1] That is, our separate measures of prior indirect experience of evaluating good and bad examples of others are not dependent on the value of the rating made by the individual, but rather on the aggregate (ex-post) score of the design being rated.

*Control Variables.* Our dependent variable *Score* derives from individual consumer ratings and the number of scores submitted varies across designs. To account for the fact that designs with fewer ratings may provide a less robust measure of performance and may be systematically higher as they are more easily influenced, we control for the count of ratings a design received (*ScoreCount*).

Consistent with the learning curve literature, we define learning to be reflected by the increase in current design performance as prior direct and indirect experience increase. However, we need to distinguish between learning from participation in our setting and changes in performance scores that may occur due to other factors (Argote & Epple, 1990; Thornton & Thompson, 2001). Hence, we include a control variable *Tenure*, which captures the number of years a designer has been active on the Threadless

---

[1] Using different cutoff values to specifying the good/bad split does not substantially change our results (e.g., mean vs. median split).

[2] We also estimated models in which we specified first-order autoregressive AR(1) as the within-subject covariance structure for repeated measures in a default error structure where subjects are assumed to be independent. Results are consistent between the different specifications.

[3] The dataset actually contains 179,794 design submissions of 56,612 individuals all of which are

platform. *Tenure* effectively also captures the general passage of time and thus incorporates possible experience accumulated outside the organization through training, education, or from other design activity. It is also important to control for other variables that could impact performance ratings such as the size of prize money offered in a competition (which may affect designers' effort), the level of competition in a given week, specific competition themes, and general changes in trend. We control for observed and unobserved competition characteristics through competition fixed effects. We constructed a competition indicator, which consists either of a unique competition ID for all one-off competitions with a specific theme or a competition ID and week indicator for the regular weekly competitions that do not have a specific theme. Thus, competition fixed effects effectively control for the competition theme, prize, and other temporal effects.

**Empirical model**

The standard model to describe learning curves is given by $AC = aExp^{-b}$ where $AC$ is the average cost of the last unit produced, *a* is the average cost of the first unit produced, *K* is the cumulative production (i.e., experience), and –*b* is the learning rate. While in the original work on learning curves the outcome measure of interest is average cost of last unit produced (e.g. Argote & Epple, 1990; Wiersma, 2007), completion time (Boh et al., 2007) or effort (Reagans, Argote, & Brooks, 2005), learning curves can also reflect more general performance outcomes such as game score (Schilling et al., 2003). In our case, we expect positive coefficients in the case of (positive) learning. The empirical equivalent of this model in this research setting to explore direct and indirect learning rates is given in Equation (1):

$$\ln Score_{it} = \alpha_i + \beta_1 \ln Exp_{it-1} + \beta_2 \ln Votes_{it-1} + \epsilon_{it} \quad (1)$$

$Score_{it}$ is the design performance of the *t*-th submission of designer *i*. $Exp_{it-1}$ is cumulative measure of all prior direct experience and $Votes_{it-1}$ is indirect experience. $\beta_1$ in Equation 1 is the learning rate from direct experience while $\beta_2$ is the learning rate from indirect experience. To test for complementarities and after accounting for control variables, competition and designer effects we estimate the following full model:

$$\ln Score_{it} = \alpha_i + \beta_1 \ln Exp_{it-1} + \beta_2 \ln Votes_{it-1} + \beta_3 \ln Exp_{it-1} * \ln Votes_{it-1} + \beta_{4-5} Controls + \beta_{6-610} Competition + \epsilon_{it} \tag{2}$$

H1 of learning from direct experience is supported if $\beta_1$ is positive and significant and H2a of learning from indirect experience is supported if $\beta_2$ is positive and significant. H2b of a complementary effect on improving design performance is supported if the coefficient $\beta_3$ of the interaction term is positive. To test H3a, if indirect experience from evaluating high- or low-performance designs is more strongly associated with improvements in design performance, we compare the coefficients $\beta_{2\,(Good)}$ and $\beta_{2\,(Bad)}$ which we test in a separate model. If $\beta_{2\,(Good)}$ is larger than $\beta_{2\,(Bad)}$, this would provide support to the hypothesis that indirect experience from evaluating high-performance designs is more important than indirect experience from evaluating low-performance designs. The hypothesis of complementarity between evaluating high- and low-performing designs is supported by a positive coefficient of the interaction term.

Our data have a crossed multilevel structure: design submissions are nested within individuals, and individuals submit to one or more competitions. We specify competitions as fixed effects and designers as random effects. Fixed effects are appropriate if the levels in the study represent all possible levels of the factor about which inferences are made, which is the case in our data, as it includes all competitions that took place during the ten-year period. A fixed-effects formulation assumes that estimating different intercept terms for each unit can best capture differences across units. This approach is appropriate if units differ in their average level of the outcome measure, which is likely in our case as competitions vary in popularity, difficulty, and prize money, which probably affects the quality of designs being submitted. A random-effects specification, on the other hand, assumes that levels of the factor used in the study represent a random sample of a larger set of potential levels (Greene, 2003). Since we wish to generalize the learning among individuals in our sample to other individuals, we define the individual level as random.

**Estimation**

Given the crossed panel nature of our data, we adopt a multilevel estimation approach (Gelman & Hill, 2006) which allows us to model fixed and random effects, accommodate the nesting of repeated measures within designers and competitions, and account for autocorrelation and unequal spacing of observations in time. We estimate the model using linear mixed model fit by maximum likelihood. We used the **lme4** (Bates, Maechler, Bolker, & Walker, 2015) and **nlme** (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2015) packages available in **R** (R Core Team, 2015) and full maximum likelihood which yields log-likelihood numbers that can be used to evaluate the incremental fit across nested models. We obtain similar results when using restricted maximum likelihood instead.

Multilevel modeling is the recommended approach for hierarchical data with multiple nested and crossed levels (Gelman & Hill, 2006). Furthermore, it affords considerable flexibility to deal with varying numbers of observations per individual and competition. We conducted a number of specification checks for the models, including specifying the competition level as random instead of fixed, and the individual level as fixed instead of random. Results are consistent with our original specifications, although our original specifications yield the best model fit. To control for serial correlation among designs submitted close in time, we estimate a model including an AR(1) covariance structure as an additional robustness test. Since design submissions are not equally spaced in time, we use a generalized AR(1) structure which allows unequal spacing between observations. Such a spatial power function models the covariance between two observations at $t_1$ and $t_2$ as $cov(y_{t1}, y_{t2}) = \sigma^2 \rho^{|t1-t2|}$, where $\rho$ is an autoregressive parameter and assumed to satisfy $|\rho| < 1$ and $\sigma^2 \rho$ is the overall variance.[2]

[2] We also estimated models in which we specified first-order autoregressive AR(1) as the within-subject covariance structure for repeated measures in a default error structure where subjects are assumed to be independent. Results are consistent between the different specifications.

## RESULTS

### Descriptive statistics

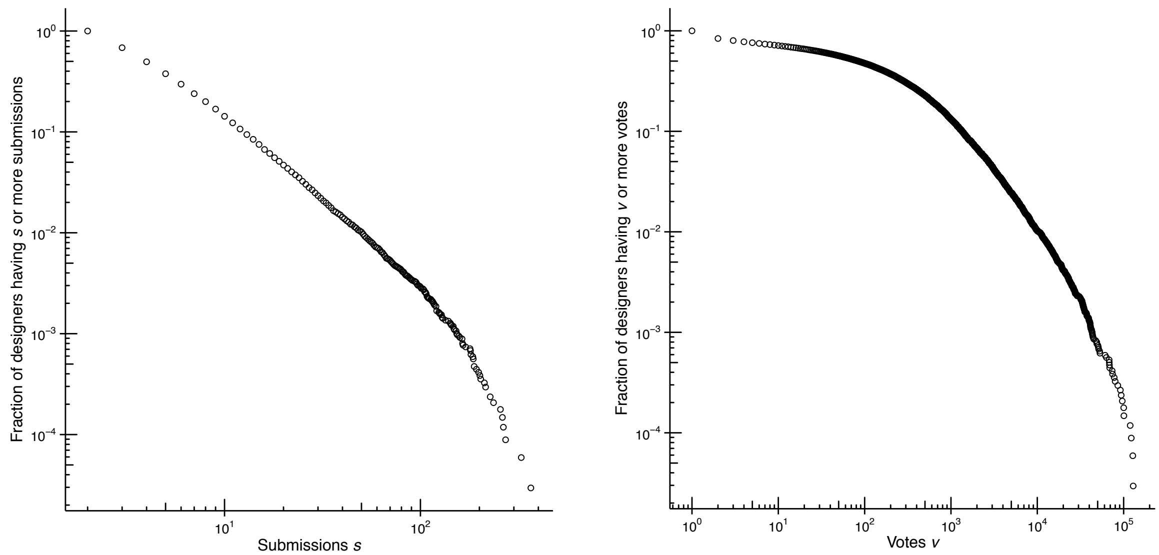
Our unbalanced, crossed panel dataset comprises 33,813 individuals with varying panel length ranging from 2-359 design submissions for a total of 136,989.[3] The panel covers a period of ten years from 2001 through 2011. Prior experience is the cumulative sum of all submissions made by an individual. The dependent variable *Score* is the aggregate of a total of 149 million ratings with an average of 851 ratings per design submission (SD=657.5). Rating is anonymous and each individual is allowed only one rating per design. Table 1 shows descriptive statistics and correlations among key variables. The average tenure of individuals making design submissions is ten months (.83 years) and a maximum of almost nine years.

**Table 1 Descriptive Statistics.**

| | Mean | SD | Min | Max | ( 1 ) | ( 2 ) | ( 3 ) | ( 4 ) |
|---|---|---|---|---|---|---|---|---|
| $Score_{it}$ ( 1 ) | 1.91 | 0.51 | 0.25 | 5.00 | | | | |
| $ScoreCount_{it}$ ( 2 ) | 850.61 | 657.46 | 1.00 | 3791.00 | 0.30 | | | |
| $Exp_{it-1}$ ( 3 ) | 16.48 | 30.84 | 1.00 | 365.00 | 0.31 | 0.06 | | |
| $Votes_{it-1}$ ( 4 ) | 2080.83 | 6498.96 | 0.00 | 128864.00 | 0.24 | 0.05 | 0.44 | |
| $Tenure_{it}$ ( 5 ) | 0.83 | 1.13 | 0.00 | 8.84 | 0.30 | 0.02 | 0.44 | 0.37 |

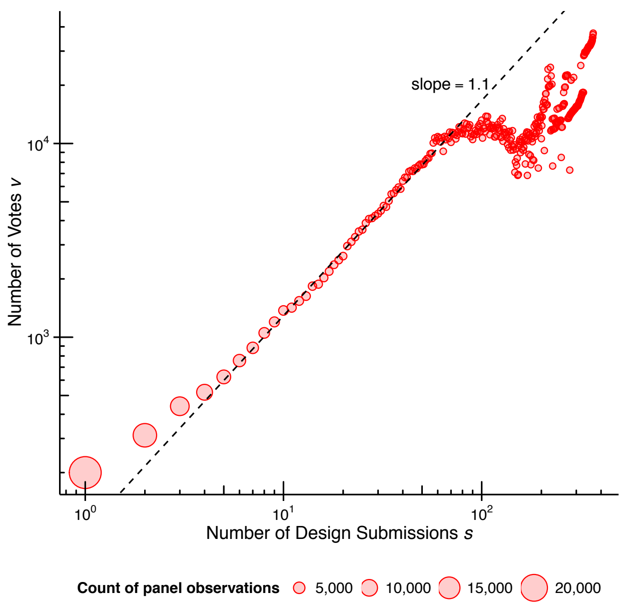
[3] The dataset actually contains 179,794 design submissions of 56,612 individuals all of which are included in the computation of key variables used in this study. However, the notion of learning is only meaningful for individuals who made at least two design submissions. Hence, the number of observations included in the panel regression includes only those individuals that made at least two design submissions.

**Figure 1 Cumulative distribution of (A) direct experience and (B) indirect experience.**




First, we begin by investigating overall patterns in making design submissions and ratings. The panel in Figure 1 shows the distribution of prior direct and indirect experience. Both follow a heavy-tail distribution. Next, we investigated if individuals' engagement in design evaluations is stable over time or whether there are any discernable patterns in behavior change. Figure 2 plots individuals' number of submissions against their number of ratings on a log-log scale. The plot shows how many ratings individuals submit on average between making design submissions over the full spectrum of individual careers. If the pattern in the data changes that would imply that the rating pattern between submissions changes while a linear pattern would imply that the number of ratings an individual casts between design submissions remains stable. From the plot we see that early in their careers (up to the $5^{th}$ design submission), individuals cast on average about 100 ratings between each design submission. Then their voting increases such that they cast about 135 ratings between each design submission. This pattern remains almost perfectly stable between the $5^{th}$ and the $50^{th}$ design submission. Among the very experienced individuals—those making their 50th or later submission—the number of ratings cast between design submissions decreases slightly to about 127. Generally, at this high level, the pattern becomes less reliable and more noisy due to fewer numbers of observations. In summary, we conclude that the pattern between design submissions and voting does not change significantly over the career of individuals and stays relatively stable at around 114 ratings between submissions across the whole dataset.

**Figure 2 Submissions versus ratings on log-log scale. Rating behavior is mostly stable between the 5th and 50th design submission (slope=1.1; translating to about 135 cast ratings between each design submission).**




## Main regression results

Table 2 presents the estimation results. We first established that there is significant variation in submission scores to be explained by estimating only an intercept term (Model 1). There is more variation in scores between submissions than there is variation between individuals. Model 2 adds controls and the direct and indirect experience variables, Model 3 adds squared terms of the experience variables, Model 4 adds an interaction term between direct and indirect experience, Model 5 replaces the overall indirect experience with separate measures for indirect experience from evaluating high- and low-performance designs, Model 6 adds an interaction term between the two measures of indirect experience, and Model 7 allows for autocorrelation to control for unequal spacing between design submissions (i.e., it uses an AR(1) correlation matrix).

**Table 2 Effects of prior direct and indirect experience on outcome (Dependent Variable: *lnScore*).**

| Dependent Variable: | *lnScore* | | | | | | |
|---|---|---|---|---|---|---|---|
| | Model 1 Baseline | Model 2 Experience | Model 3 Squared Terms | Model 4 Direct-Indirect Complement. | Model 5 Good vs. Bad | Model 6 Good-Bad Complement. | Model 7 Unequal Spacing |
| *Intercept* | 0.868*** (0.163) | 1.119*** (0.133) | 1.112*** (0.133) | 1.154*** (0.133) | 1.120*** (0.133) | 1.153*** (0.133) | 1.119*** (0.134) |
| *lnExp* | | 0.004*** (0.001) | −0.013*** (0.001) | 0.002** (0.001) | 0.004*** (0.001) | 0.001* (0.001) | 0.004*** (0.001) |
| *lnVotes* | | 0.010*** (0.000) | 0.005*** (0.001) | 0.013*** (0.000) | | | 0.010*** (0.000) |
| $lnExp^2$ | | | 0.005*** (0.000) | | | | |
| $lnVotes^2$ | | | 0.001*** (0.000) | | | | |
| *lnExp* × *lnVotes* | | | | 0.003*** (0.000) | | | |
| $lnVotes_{Good}$ | | | | | 0.017*** (0.001) | 0.024*** (0.001) | |
| $lnVotes_{Bad}$ | | | | | −0.007*** (0.001) | −0.010*** (0.001) | |
| $lnVotes_{Good}$ × $lnVotes_{Bad}$ | | | | | | 0.002*** (0.000) | |
| *Tenure* | No | Yes | Yes | Yes | Yes | Yes | Yes |
| *ScoreCount* | No | Yes | Yes | Yes | Yes | Yes | Yes |
| *Competition* | Fixed | Fixed | Fixed | Fixed | Fixed | Fixed | Fixed |
| *Designer* | Random | Random | Random | Random | Random | Random | Random |
| Log Likelihood | 13137.699 | 40248.207 | 40435.414 | 40416.161 | 40320.697 | 40449.323 | 40248.207 |
| Variance | | | | | | | |
| Designer (Intercept) | 0.027 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 | 0.013 |
| Residual | 0.038 | 0.026 | 0.026 | 0.026 | 0.026 | 0.026 | 0.027 |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

The results indicated that an individual's current design performance increases with increasing prior direct experience ($\beta = .004;\ p < .001$), supporting H1, and indirect experience ($\beta = .010;\ p < .001$), supporting H2a. We provide additional analyses to better interpret the size of the coefficients in a simulation study that follows. To test whether direct or indirect experience is more effective, we used a Wald chi-square test to compare the two regression coefficients. The test confirmed that indirect experience has a stronger effect on individuals' learning ($\chi^2 = 39.82;\ p < .001$). Given the logged values in our model, the coefficients can be interpreted as percentage changes: doubling an individual's indirect experience from voting results in a larger expected increase in that individual's current design performance than doubling that individual's direct experience. However, it is important to bear in mind that design submissions and cast ratings operate on different orders of magnitude with roughly a rate of 1:100 (see Figure 2). For example, an average individual with a prior direct experience of five design submissions will usually have cast about 500 ratings. So in order to double his/her direct experience that individual would need to submit another five designs but would have to cast another 500 ratings to double indirect experience.

We found statistically significant squared terms of both direct and indirect experience. This suggests that the learning curve for both types of experience is non-linear (we investigate the exact shape of the learning curves in more detail below). To test whether direct experience and indirect experience have a complementary effect on improving an individual's current design performance, we included an interaction term between the two. We found a positive and statistically significant effect ($\beta = .003;\ p < .001$), which supports H2b, suggesting a complementarity between direct and indirect experience.

With regard to learning from indirect experience from evaluating high- and low-performance designs, we found a statistically significant and positive coefficient for indirect experience from evaluating high-performance designs ($\beta = .017;\ p < .001$) and a statistically significant but negative coefficient for indirect experience from evaluating low-performance designs ($\beta = -.007;\ p < .001$). This indicates that while individuals are able to learn from evaluating good examples, they are unable to learn from bad examples, which supports H3a. Next, we evaluated the complementarity between evaluating high- and low-performing designs by introducing an interaction term of the two measures. We found a statistically significant and positive coefficient ($\beta = .002;\ p < .001$), which indicates that prior experience in evaluating others' good work can help individuals overcome their inability to learn from others' bad work, supporting H3b.

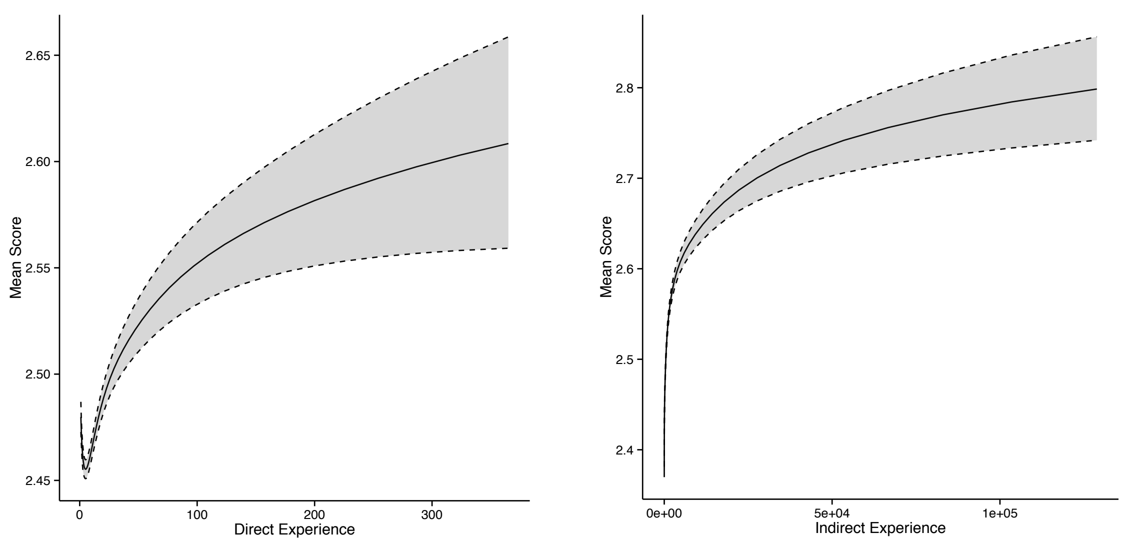
Model 7 controls for autocorrelation between submissions and the fact that observations in our panel model are not equally spaced in time. The results using this model specification are virtually identical to the specification that does not assume autocorrelation. We also tested if slopes vary in addition to intercepts, but we found that a model specification with varying intercepts only provides a better fit for the data.

**Shape of the learning curve**

Model 3 in Table 2 includes squared terms for both direct and indirect experience. The coefficients for both squared terms are statistically significant, thus establishing that the relationship is non-linear. But what exactly is the shape of the learning curve? Leveraging our extremely rich dataset of over 130,000 design submissions and long panels, we directly investigated the shape of learning curves using generalized additive models (GAM) which we fit via maximum-likelihood (James, Witten, Hastie, & Tibshirani, 2013).

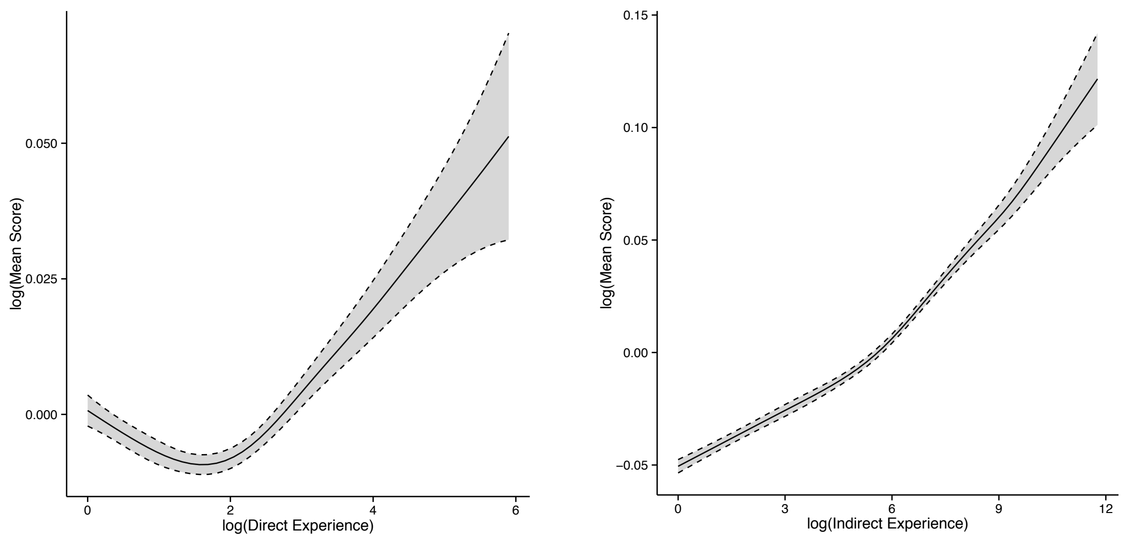
The analysis confirmed that the non-linear smooth terms are a significantly better fit for the data than a linear model ($lnExp$ has 4.56 degrees of freedom, $p < .001$; $lnVotes$ has 4.57 degrees of freedom, $p < .001$). The non-parametric GAM model explained 26.8% of deviance. Below we show two panels of the shape of the design learning curve. The first panel (Figure 3) shows the learning curve on the normal scale thus correlating design performance (on 0-5 scale) and prior experience directly. While this view is common in the learning curve literature (e.g. Argote & Epple, 1990), the shape is difficult to interpret given the wide range of values of prior experience. The second panel (Figure 4) shows the learning curve on the log-scale thus showing the actual rate of learning rather than the resulting direct improvement in design performance.

**Figure 3 Learning curve in absolute terms: how good are individuals given certain levels of experience? Absolute performance of direct experience (*Direct Experience*; left) and indirect experience (*Indirect Experience; right).***



*Note.* Models estimated using flexible GAM shown with 95% confidence interval (Dependent Variable: *lnScore*). Model based on Model (3) from Table 2. Shape is on the scale of the response variable. Plot includes overall intercept but excludes competition and desinger specific intercepts.

**Figure 4 What is the actual rate of learning? Shape of learning curves for direct Experience (left panel) and indirect experience (right panel).**



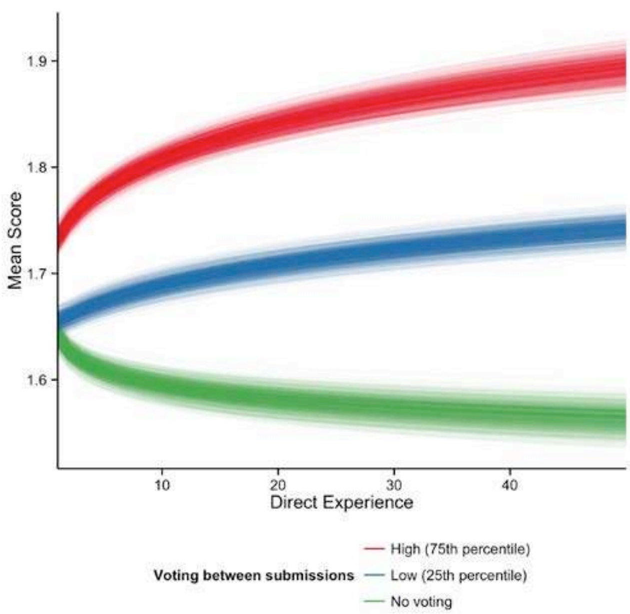

*Note.* Models estimated using flexible GAM shown with 95% confidence interval (Dependent Variable: *lnScore*). Model based on Model (3) from Table 2. Shape is on the scale of the independent variable. Plot excludes competition and desinger specific intercepts.

We interpret the second panel using the log-scale. For direct experience, the shape is downward sloping early in individuals' careers (from 1st through about the 7th submission) indicating a period of decreased performance under direct experience alone. After about the 8th submission the learning curve slopes upward and positive learning accrues. Even very late in individuals' careers, after about 150 submissions, the learning curve is positive and of the same slope indicating that even highly experienced individuals continue to learn from direct experience. For indirect experience, there does not appear to be a period of decreased performance as the curve is sloped upward even for very low levels of experience. This upward slope continues until very late in individuals' careers. We address the dip of the learning curve under direct experience in our discussion section.

**Complementarity of direct and indirect experience**

To explore the complementary effect of direct and indirect experience (our H2b), we simulated canonical behavior and plotted the resulting learning curves. We chose three canonical scenarios: an individual who does not engage in voting between design submissions at all; an individual with a moderate amount of voting activity (four evaluations between submissions which corresponds to the 25$^{th}$ percentile of voting volume); and an individual with high rating activity (140 ratings between submissions corresponding to the 75$^{th}$ percentile of rating volume). The simulation follows the approach suggested by King, Tomz, and Wittenberg (2000) and is displayed as water-color visually weighted regression (Hsiang, 2013). We simulated 1,000 instances for each of the three canonical scenarios. Results are shown in Figure 5 using the measures on the natural scale relating experience directly to performance (i.e., not log-transformed). Results showed that rather than just being additive (higher intercept), the effect is multiplicative and the slope of the curve changes with increased numbers of ratings between design submissions.

**Figure 5 Canonical Individual Behavior shows Complementary of Direct and Indirect Learning. Simulation of learning curves of three canonical individual behaviors. Bottom/Green: individual who does not engage in rating between submissions shows decreasing performance. Middle/Blue: individual who engages in low level of rating (25$^{th}$ percentile; 4 ratings between submissions) shows positive and increasing performance. Top/Red: individual who engages in high level of rating (75$^{th}$ percentile; 140 ratings between submission) shows positive and increasing performance.**



Without any voting between design submissions, the performance decreases continuously and an individual is not able to improve the performance of subsequent design submissions.

Even with just a few evaluations between design submissions, performance increases (on the log-scale this means that more and more improvement is derived from subsequent design submissions). High voting activity substantially complements direct experience such that increasing benefits derive from subsequent design submissions.

**Learning mechanisms from community-based versus firm-based evaluations**

Our analyses show significant positive learning from direct and indirect experience. However, the results raise questions regarding the underlying mechanism of learning: What is the mechanism through which individuals learn to produce more successful design submissions? What *exactly* are individuals learning? Are they learning how to produce designs to meet community-based evaluations or are they learning to cater better to the idiosyncratic style of Threadless and meet its firm-based choices? To investigate this question, we perform supplementary time-series-cross-section analyses using a binary dependent variable indicating whether a design submission has been selected for printing in a given week (Einhorn & Hogarth, 1981). This set of analyses allowed us to address the question of what role experience plays in winning a competition as opposed to receiving higher scores from community evaluations. In addition to addressing the questions about what individuals learn, this set of analyses also allows us to address the more general question of what role experience plays in winning a competition, rather than just receiving higher scores in the community evaluation. This analysis uses two separate evaluations. The evaluation score used in the main analysis is a judgment-based evaluation by the community in which designs are evaluated independently by assignment of an evaluation score (0 through 5 stars; which is then averaged across all ratings) and the second measure is a choice-based evaluation by which the organization Threadless chooses winners among the set of submissions made to the same competition.[4]

We restricted our analysis to the first time an individual wins a competition (makes a design submission that is chosen for print) and ignore observations after an individual has won for the first time. We follow the approach of Beck, Katz, and Tucker (1998) and Carter & Signorino (2010) and estimate the likelihood of winning a competition in a given week *t* using

$$\Pr(y_i = 1 | x_i, t) = \frac{1}{1 + \exp[-(x_i\beta + s(t_i)).]}. \tag{3}$$

We estimate the probability of winning conditional on an individual *i*'s time-varying measures $x_i$ which in our case are measures of direct and indirect experience and $s(t_i)$ is a simple cubic polynomial $t_i + t_i^2 + t_i^3$. We can thus investigate how important direct and indirect experience is for winning a competition in a given week by the regression coefficients. We can furthermore quantify the baseline hazard rate of winning a competition in a given week and how much more likely an individual becomes to win a competition from each additional unit of prior experience. A positive and significant coefficient for a measure of experience would indicate that the likelihood of winning a competition increases with additional experience. We would expect the intercept to be negative and quite large as the baseline rate of winning a competition is very low (only about 1-3 submissions get printed out of about 600 submissions every week). We estimate the equation using logistic regression (GLM).

To test whether individuals learn from their prior direct and indirect experience about idiosyncratic choice-based decision *above and beyond* what they learned about producing designs that receive higher judgment-based community evaluation, we estimate a second model in which we also include a time-varying measure of the highest design score based on the community judgment that an individual has been able to produce up to that point (we simply take the maximum score that an individual

[4] Results of judgment and choice decisions are not necessarily identical as judgment can be ignored at the point of choice (Einhorn & Horgath, 1981).

achieved up to that point). The intuition behind the analysis is as follows: if the coefficient for either direct or indirect experience is still significant when controlling for the community-based score that an individual has gained, then the individual must have learned something from this experience about Threadless' decision making process of selecting competition winners.

**Table 3 Effects of prior direct and indirect experience on winning a competition (Dependent Variable: Pr(*Won*=1, t)).**

| | Model 1 Experience | Model 2 Quality |
|---|---|---|
| *Intercept* | $-8.823^{***}$ | $-15.403^{***}$ |
| | (0.072) | (0.154) |
| *Exp* | $0.039^{***}$ | $0.022^{***}$ |
| | (0.001) | (0.001) |
| *Votes* | $0.000^{***}$ | $0.000^{***}$ |
| | (0.000) | (0.000) |
| *Max Score* | | $3.360^{***}$ |
| | | (0.049) |
| $t$ | $-0.019^{***}$ | $-0.036^{***}$ |
| | (0.003) | (0.003) |
| $t^2$ | $-0.000$ | $0.000^{**}$ |
| | (0.000) | (0.000) |
| $t^3$ | 0.000 | $-0.000$ |
| | (0.000) | (0.000) |
| AIC | 18297.026 | 13652.833 |
| Log Likelihood | $-9142.513$ | $-6819.416$ |
| Deviance | 18285.026 | 13638.833 |
| N | 21,799,462 | 21,799,462 |

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

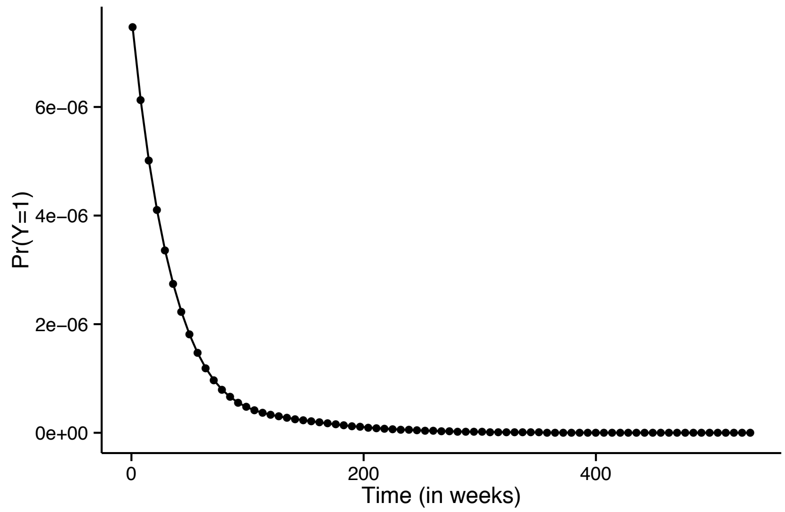
Table 3 shows the regression results for the time-series-cross-section analysis. Model 1 includes our two key measures for direct and indirect experience along with an intercept. We begin by interpreting the intercept and the baseline hazard rate of winning. As expected, the intercept is negative and quite large ($\beta = -8.823;\ p < .001$), indicating a low overall likelihood of winning. Figure 6 shows the baseline hazard rate of winning the first competition as a function of time. The hazard rate is decreasing nonlinearly and approaches zero. In other words, the chance of winning is extremely low (basically approaching zero) and in the absence of any experience from making design submissions or engaging in the evaluation of designs of others, decreases even further over time. The non-linearity of the curve results from including time polynomials. [5]

Next, we investigate the role of direct and indirect experience in winning a competition. The coefficients for direct and indirect experience are both positive and statistically significant (both $p < .001$), indicating that experience improves the odds of winning a competition over the (low) baseline hazard rate. Since the coefficients are quite small, we compute relevant quantities of interest to aid in interpretation. All else equal, direct experience from submitting one more design increases the chance of winning by 4%. All else equal, indirect experience from evaluating 100 additional designs increases the chance of winning by 0.6%.

Model 2 introduces the additional covariate capturing the quality of the best design an individual has been able to produce so far based on average community judgment. The coefficient is quite large and statistically significant ($\beta = -3.36;\ p < .001$), indicating that having demonstrated the ability to produce high-scoring designs in the past significantly increases the chance of winning a competition. More important for our analysis is, however, that the coefficients for direct and indirect experience both remain statistically significant, albeit decreasing in size. This indicates that—controlling for past high-scores captured by the community-based judgment—individuals still benefit from both direct and indirect experience. That is, they learn about specific idiosyncrasies of the decision-making process by which winners are chosen.

[5] Instead of using time, time squared, and time cubed, we also test an alternative model specification which models time as a penalized cubic regression spline—the method originally proposed by Beck, Katz, and Tucker (1998)—which we estimate using a GAM model. We find almost identical coefficients than those reported in Table 3. The spline is estimated to have 3.8 knots and a test confirms that the non-linear spline is a significantly better fit for the data than a linear model.

**Figure 6 Baseline hazard rate of winning the first competition over time (from Model 2).**




## DISCUSSION

Our large scale dataset allowed us to directly explore learning in creative design work. We have five main results, which we briefly summarize here and then expand upon in the following sections. First, we establish that even in the realm of creative design work, individual learning follows a learning curve associated with both direct and indirect experience. Second, while learning curves reported in the literature demonstrate continual positive improvement, we find that in the absence of vicarious learning, individuals surprisingly experience “design myopia,” where individuals initially see decreased performance with increased direct experience. Third, we find strong complementarities between direct and indirect experience that can help individuals overcome this “design myopia.” Fourth, we find that individuals are unable to learn from other’s bad work, though experience from evaluating others’ good can help individuals overcome their inability to learn from others’ bad examples. Finally, we find that direct and indirect experience helps individuals not only to gain high ratings from consumers but also to understand the “black box” of how designs are chosen by firms for production. We next expand on these findings.

### Quantifying learning and design myopia

Our empirical approach allows us to quantify learning in the context of creative design work, a domain where there is even debate whether learning can occur. One prior study of crowdsourced product innovation on the Dell IdeaStorm.com platform, for example, has found no learning effects (Huang et al., 2014), and this difference in findings may lie in part due to the nature of the task between design work and crowdsourcing product innovations. In creative design work an individual is able to focus their creative effort, such as within the realm of T-shirt design or a car design, where repeated submission of designs builds upon past experience in developing design skills and understanding customer preferences; in contrast, in crowdsourced product innovation one is rewarded for branching out and seeking out novel and disparate solutions, with little opportunity for cumulative learning.

Not only do we demonstrate learning from experience, we extend past studies of creative work that looked at relative rates of learning (Gino et al., 2010) to detail the rate of progress over time. In fact, due to the challenge of collecting large scale data on indirect experience, learning curves typically have been done on the basis of direct experience only (e.g. Argote & Epple, 1990; Newell & Rosenbloom, 1981). Learning curves on the basis of indirect experience are rare, and our study illuminates how both direct and indirect experience contribute to learning, with diminishing returns, over time.

Past studies of learning curves in routine work were focused on the acquisition of specific skills, such as in an manufacturing context. In design work, there are specific technical skills (such as graphic design skills) that individuals acquire and develop, but there is additionally the need to develop an appreciation for and respond to what is valued by consumers. While our empirical setting does not allow us to disentangle the degree to which learning was centered in skills versus appreciation for consumer preferences, we provide important insights into underlying mechanisms regarding learning from others' good and bad work, and that individuals can learn about an idiosyncratic decision-making process above and beyond what they learn from the judgment-based community evaluations.

We also uncover an unexpected aspect of learning creative design work through direct experience. Existing theories of learning typically assume progress at the outset of gaining direct experience. Though the learning curve of design broadly mirrors that of other contexts, an empirical difference was that in the absence of indirect learning, there was an initial period of decreasing performance, what we have termed "design myopia." Empirically, this finding is surprising given the lack of such a dip in other learning curves, but an examination of the design process helps us to understand what is likely behind this phenomenon. Myopic processes have been seen in other areas of organizational life, where individuals fail to learn due to a narrowing of their focus, often on the basis of discounting "distant" information (Levinthal & March, 1993).The social nature of design evaluation can contribute to a delay in gaining improvement in design quality. A good design exhibits fit not only among its elements but also between its elements and the consumer preferences and the environment more broadly (Alexander, 1964; Clark, 1985). Initial direct experience relies on feedback in the form of evaluations of design competitions, but these evaluations will be ambiguous as to the cause of the quality assessment and underlying customer preferences. Resolving the ambiguity will take further direct experience, as one experiments with repeated practice and with different ways to read a market.

While the ambiguous nature of customer preference would explain a delay to positive learning, by itself it does not explain *negative* learning—actual decreases in performance with each increment of experience—in the earliest stages absent any vicarious learning. We posit an explanation for this negative learning lies with myopia in the absence of learning from others, coupled with individuals drawing from a "back-log" of design ideas. In creative design work there may be somewhat similar goals between assignments. For example, a car designer may design a succession of mid-size cars, a toy designer may have several assignments in a similar market, and an apparel designer may be designing for a specific style. A designer coming to a creative design competition may have a "back-log" of design ideas that they would try out, putting forth their most promising creative idea first, what they view as their next-most creative idea second, and so-on, without attending to vicarious learning opportunities from observing others.

A practical implication of this empirical finding is that by shifting from their own catalog of what they consider "good" designs to observing and evaluating the work of others, the individual would be able to more quickly improve. The evidence for design myopia likewise provides avenues for further research on how initial feedback from direct experience may lead to decreases in performance across different design domains; the depth and length of this effect can be characterized by comparative work.

### Vicarious learning from good versus bad examples

We shed light on under what conditions individuals fail to learn from indirect experience. Specifically, we find that individuals fail to learn from bad examples. On the face of things, this result could be puzzling on two fronts. First, prior work across domains has highlighted the benefits of learning from failure (Firestein, 2015; Madsen & Desai, 2010; Sitkin, 1992). Second, additional work has focused on the benefits of learning from *others'* failures, too (Edmondson, 1999; KC et al., 2013). In managerial settings, learning from failure is important because we risk making poorly-formed decisions if we neglect learning from others' failures rather than focusing merely on success (Denrell, 2003).

The unique aspects of creative design work help to explain why negative examples may not play the same role as failures in other work. In most work tasks, failures often contain critical and focused information about what to avoid: for example, how a surgical procedure can go wrong (KC et al., 2013) or how a space mission may be threatened (Madsen & Desai, 2010); in contrast, the new information contained in successes in such contexts is relatively small; observing another's successful surgical procedure may yield some new insights on how to further improve efficiency, but not as much information on what to avoid due to a procedure going awry. In the context of creative design work, the opposite dynamic is in effect. An example of good design has been able to integrate an array of difficult to discern consumer preferences and produce a well-received design. It is difficult to learn how to design and how to meet consumer needs, and so those who produce "good" designs have successfully integrated a number of factors into their work. In contrast, a bad example has failed to understand and adapt to consumer preferences, but there are many ways to make these mistakes, and the mere evidence of a failure yields little insight into how to improve one's own work.

At the time of evaluating designs for themselves, in our empirical situation an individual does not know how others have rated the design more broadly, as average scores were not available until voting had ended. This mimics common instances in creative design work where one is exposed to a range of designs, not all of which are officially judged. Merely by being exposed to more good designs, an individual is exposed to positive cognitive templates for designs. Cognitive templates are learning heuristics that help individuals to create ways to develop courses of action (Bingham & Kahl, 2013; Gick & Holyoak, 1983). By being exposed to a range of good examples, individuals have templates of such things as what designs might be most likely to gain acceptance, what colors may be values, or what fonts might be in style.

A related finding was that increased experience with good examples can overcome the relative challenge of working with bad examples. The theoretical basis for explaining this phenomenon may best lie in how a combination of good and bad examples together helps to define the "design space" (Erat & Kavadias, 2008) from which an individual can draw inspiration. Design spaces are considered to encompass the range of possible choices a design may make under a given set of constraints (Erat & Kavadias, 2008). By being exposed to a combination of good and bad examples, an individual is able to better understand the dimensions of this design space and more intelligently make decisions about their own designs.

**Learning to improve above and beyond community-based evaluations**

While the results discussed above provide important insights into the mechanisms of learning when considering community-based judgment as the source of evaluation broadly, we also address how experience played a role in meeting firm-based choice evaluations. It was quite conceivable that individuals learned about how to cater to members of an online community without learning to meet the decision-criteria for being chosen by the firm to win a competition. Our analyses of the likelihood of winning a design competition reveals that individuals learn from their experiences above and beyond how to produce designs that achieve high community evaluations. This is quite surprising, since Threadless picks winners internally and the process is entirely up to them and thus difficult to observe. In effect, all submitted designs enter a "black box" and a winner emerges. Our analysis demonstrates that individuals are able to learn how this black box functions and learn to produce designs that stand higher chances of winning. While recent research has called attention to how similar community-based judgments can be to expert choices in crowdsourced funding efforts (Mollick & Nanda, 2015), our study points to how different aspects may be learned in the realm of creative design work. In creative design work, we find evidence of learning not only to gauge the preferences of a broad community of fellow consumers but also of the underlying processes by which choices are made by firms. In design work more broadly, individuals not only tailor their designs to the public, they may also seek to win awards and honorable mentions from experts and critics in their field (e.g. as catalogued in Fiell & Fiell, 2005), and our study demonstrates that both aspects are learned with experience.

**CONCLUSION**

Creative design work is responsible for bringing about many product and cultural innovations, yet we have known little about how creative design is improved over time—especially when compared with the vast research done on learning in other contexts. Our study drew on an empirical setting well-suited to probe the mechanisms of learning in creative design work. "Design myopia" and the relative value of good versus bad designs were shown to be important factors shaping learning in creative design work. This study provides a basis for additional work that can continue to examine how this fundamental activity within so many organizations is developed over time, how the design learning curve may change in different contexts, and additional factors that accelerate or impede learning. Creative design work is an activity by which individuals are able to change the material and aesthetic nature of our society, and such a subject is worthy of continued large-scale empirical examination.

**REFERENCES**


Alexander, C. 1964. ***Notes on the synthesis of form***. Cambridge, Massachusetts: Harvard University Press.

Argote, L., & Epple, D. 1990. Learning curves in manufacturing. ***Science***, 247(4945): 920-924.

Argote, L., & Miron-Spektor, E. 2011. Organizational learning: From experience to knowledge. ***Organization science***, 22(5): 1123-1137.

Audia, P. G., & Goncalo, J. A. 2007. Past success and creativity over time: A study of inventors in the hard disk drive industry. ***Management Science***, 53(1): 1-15.

Bandura, A. 1977. ***Social learning theory***. Englewood Cliffs, NJ: Prentice-Hall.

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. 2015. Fitting Linear Mixed-Effects Models using {lme4}, ***Retreived from: http://arxiv.org/abs/1406.5823***. Retreived from: http://arxiv.org/abs/1406.5823.

Beck, N., Katz, J. N., & Tucker, R. 1998. Taking time seriously: Time-series-cross-section analysis with a binary dependent variable. ***American Journal of Political Science***: 1260-1288.

Bingham, C. B., & Kahl, S. J. 2013. The Process of Schema Emergence: Assimilation, Deconstruction, Unitization and the Plurality of Analogies. ***Academy of Management Journal***, 56: 14-34.

Bloch, P. H. 1995. Seeking the ideal form: Product design and consumer response. ***Journal of Marketing***, 59: 16-29.

Boh, W. F., Slaughter, S. A., & Espinosa, J. A. 2007. Learning from Experience in Software Development: A Multilevel Analysis. ***Management Science***, 53(8): 1315-1331.

Boudreau, K. J., & Lakhani, K. R. 2009. How to manage outside innovation. ***MIT Sloan Management Review***, 50(4): 69-76.

Bresman, H. 2010. External Learning Activities and Team Performance: A Multimethod Field Study. ***Organization Science***, 21(1): 81-96.

Carter, D. B., & Signorino, C. S. 2010. Back to the future: Modeling time dependence in binary data. ***Political Analysis***: mpq013.

Clark, K. B. 1985. The interaction of design hierarchies and market concepts in technological evolution. ***Research Policy***, 14(5): 235-251

Denrell, J. 2003. Vicarious Learning, Undersampling of Failure, and the Myths of Management. ***Organization Science***, 14(3): 227-243.

Edmondson, A. C. 1999. Psychological safety and learning behavior in work teams. ***Administrative Science Quarterly***, 44(2): 350-383.

Einhorn, H. J., & Hogarth, R. M. 1981. Behavioral decision theory: Processes of judgment and choice. ***Annual Review of Psychology***, 32: 53-58.

Eisenman, M. 2013. Understanding aesthetic innovation in the context of technological evolution. ***Academy of Management Review***, 38(3): 332-351.

Erat, S., & Kavadias, S. 2008. Sequential Testing of Product Designs: Implications for Learning. ***Management Science***, 54(5): 956-968.

Fiell, C., & Fiell, P. (Eds.). 2005. ***Designing the 21st century***. Koln, Germany: Taschen.

Firestein, S. 2015. ***Failure: Why science is so successful***. Oxford, UK: Oxford University Press.

Gelman, A., & Hill, J. 2006. ***Data analysis using regression and multilevel/hierarchical models***. Cambridge, UK: Cambridge University Press.

Gick, M. L., & Holyoak, K. J. 1983. Schema induction and analogical transfer. ***Cognitive psychology***, 15(1): 1-38.
Gino, F., Argote, L., Miron-Spektor, E., & Todorova, G. 2010. First, get your feet wet: The effects of learning from direct and indirect experience on team creativity. ***Organizational Behavior and Human Decision Processes***, 111(2): 102-115.
Greene, W. H. 2003. ***Econometric analysis*** (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
Hargadon, A., & Douglas, Y. 2001. When innovations meet institutions: Edison and the design of the electric light. ***Administrative Science Quarterly***, 46: 476-501.
Hargadon, A., & Sutton, R. I. 1997. Technology brokering and innovation in a product development firm. ***Administrative Science Quarterly***, 42(4): 716-749.
Harrison, S. H., & Rouse, E. D. 2015. An Inductive Study of Feedback Interactions over the Course of Creative Projects. ***Academy of Management Journal***, 58(2): 375-404.
Harvey, S. 2014. Creative Synthesis: Exploring the Process of Extraordinary Group Creativity. ***Academy of Management Review***, 39(3): 324-343.
Hsiang, S. M. 2013. Visually-weighted regression. In P. University (Ed.), ***Princeton University Working Paper***. Available at: http://bit.ly/Ld8WxP.
Huang, Y., Singh, P. V., & Srinivasan, K. 2014. Crowdsourcing New Product Ideas Under Consumer Learning. ***Management Science***, 60(9): 2138-2159.
James, G., Witten, D., Hastie, T., & Tibshirani, R. 2013. ***An introduction to statistical learning with applications in R***. New York: Springer.
Kane, G. C., Johnson, J., & Majchrzak, A. 2014. Emergent life cycle: The tension between knowledge change and knowledge retention in open online coproduction communities. ***Management Science***, 60(12): 3026-3048.
KC, D., Staats, B. R., & Gino, F. 2013. Learning from My Success and from Others' Failure: Evidence from Minimally Invasive Cardiac Surgery. ***Management Science***, 59(11): 2435-2449.
King, G., Tomz, M., & Wittenberg, J. 2000. Making the most of statistical analyses: Improving interpretation and presentation. ***American Journal of Political Science***, 44(2): 347-361.
Levinthal, D. A., & March, J. G. 1993. The myopia of learning. ***Strategic Management Journal***: 95-112.
Madsen, P. M., & Desai, V. 2010. Failing to Learn? The Effects of Failure and Success on Organizational Learning in the Global Orbital Launch Vehicle Industry. ***Academy of Management Journal***, 53(3): 451-476.
Mollick, E., & Nanda, R. 2015. Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts. ***Management Science***, Articles in Advance(0): 1-12.
Newell, A., & Rosenbloom, P. S. 1981. Mechanisms of skill acquisition and the law of practice. In J. R. Anderson (Ed.), ***Cognitive skills and their acquisition***: 1-55. Hillsdale, NJ: Erlbaum.
Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. 2015. Linear and non-linear mixed effects models. Retrieved from http://cran.r-project.org/package=nlme.
Pisano, G. P., Bohmer, R. M. J., & Edmondson, A. C. 2001. Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimally Invasive Cardiac Surgery. ***Management Science***, 47(6): 752-768.
R Core Team. 2015. R: A language and environment for statistical computing. Vienna, Austria: Retrieved from http://www.r-project.org/.

Ravasi, D., & Lojacono, G. 2005. Managing design and designers for strategic renewal. ***Long Range Planning***, 38: 51-77.
Reagans, R., Argote, L., & Brooks, D. 2005. Individual Experience and Experience Working Together: Predicting Learning Rates from Knowing Who Knows What and Knowing How to Work Together. ***Management Science***, 51(6): 869-881.
Rindova, V., Dalpiaz, E., & Ravasi, D. 2011. A cultural quest: A study of organizational use of new cultural resources in strategy formation. ***Organization Science***, 22(2): 413-431.
Rindova, V. P., & Petkova, A. P. 2007. When is a new thing a good thing?  Technological change, product form design, and perceptions of value for product innovations. ***Organization Science***, 18(2): 217-232.
Schilling, M. A., Vidal, P., Ployhart, R. E., & Marangoni, A. 2003. Learning by Doing Something Else: Variation, Relatedness, and the Learning Curve. ***Management Science***, 49(1): 39-56.
Seidel, V. P., & Langner, B. 2015. Using an online community for vehicle design: Project variety and motivations to participate. ***Industrial and Corporate Change***, 24(3): 635-653.
Seifert, M., Siemsen, E., Hadida, A. L., & Eisingerich, A. B. 2015. Effective judgmental forecasting in the context of fashion products. ***Journal of Operations Management***, 36(0): 33-45.
Simon, H. A. 1996. ***The sciences of the artificial*** (3rd ed.). Cambridge, Mass.: MIT Press.
Sitkin, S. B. 1992. Learning Through Failure: The Strategy of Small Losses. In B. M. Staw, & L. L. Cummings (Eds.), ***Research in Organizational Behavior***, Vol. 14. Greenwich, CT: JAI Press.
Srinivasan, R., Haunschild, P., & Grewal, R. 2007. Vicarious learning in new product introductions in the early years of a converging market. ***Management Science***, 53(1): 16-28.
Thornton, R. A., & Thompson, P. 2001. Learning from experience and learning from others: An exploration of learning and spillovers in wartime shipbuilding. ***American Economic Review***, 91(5): 1350-1368.
Uluoğlu, B. 2000. Design knowledge communicated in studio critiques. ***Design Studies***, 21(1): 33-58.
von Krogh, G., Haefliger, S., Spaeth, S., & Wallin, M. W. 2012. Carrots and rainbows: Motivation and social practice in open source software development. ***MIS Quarterly***, 36(2): 649-676.
Wiersma, E. 2007. Conditions That Shape the Learning Curve: Factors That Increase the Ability and Opportunity to Learn. ***Management Science***, 53(12): 1903-1915.
Wijnberg, N. M., & Gemser, G. 2000. Adding Value to Innovation: Impressionism and the Transformation of the Selection System in Visual Arts. ***Organization Science***, 11(3): 323-329.
Wright, T. P. 1936. Factors affecting the cost of airplanes. ***Journal of the Aeronautical Sciences***, 3(4): 122-128.