On Styles in Product Design: An Analysis of U.S. Design Patents

Tian Heong Chan, a Jürgen Mihm, b Manuel E. Sosa c

a Information Systems and Operations Management, Goizueta Business School, Emory University, Atlanta, Georgia 30322; b Technology and Operations Management, INSEAD, 77300 Fontainebleau, France; c Technology and Operations Management, INSEAD, Singapore 138676

Contact: tian.chan@emory.edu (THC); jurgen.mihm@insead.edu (JM); manuel.sosa@insead.edu (MES)

Received: April 11, 2014
Accepted: July 14, 2016
Published Online in Articles in Advance: February 21, 2017

https://doi.org/10.1287/mnsc.2016.2653

Abstract. Products combine function and form. This paper focuses on product form. We combine state-of-the-art clustering techniques with experimental validation to identify styles (groupings of new product designs of similar form) among the more than 350,000 U.S. design patents granted from 1977 through 2010. Thus we compile, for the first time, a rich data set of styles that can serve as an empirical platform for a rigorous study of the role played by product form in new product development. Building on this platform, we analyze the determinants of “style turbulence”—the year-to-year unpredictability of changes in a style’s prevalence. We find that (i) style turbulence follows a U-shaped relationship with respect to function turbulence (the turbulence of product functions associated with a given style), and (ii) style turbulence increases over time. We discuss the implications of these findings for managing design in new product development.

History: Accepted by David Hsu, entrepreneurship and innovation.
Funding: This research has been partially funded by INSEAD’s Heinrich and Esther Baumann-Steiner Endowed Fund for Creativity and Business.
Supplemental Material: Data and the e-companion are available at https://doi.org/10.1287/mnsc.2016.2653.

Keywords: product form • cluster analysis • experimental validation • design patents • industry turbulence • design innovation

1. Introduction

Previous work in innovation and technology management has shaped our understanding of what drives the evolution of products. These studies have examined how discontinuities in technological advances redefine technological frontiers, initiate entirely new sets of product categories, and disrupt established players (Anderson and Tushman 1990, Baldwin and Clark 2000, Henderson and Clark 1990, Tushman and Anderson 1986, Utterback 1996, Utterback and Abernathy 1975). This large body of work rests on the assumption that technology, and hence function (how a product works), is the major source of change in product evolution.

However, a product combines elements of both function and form, or how it looks (Alexander 1964, Ulrich 2011). Indeed, the dynamics of such industries as clothing, furniture, kitchenware, and (more recently) computing seem to be driven as much by changes in form as in function (Dell’Era et al. 2010). It is widely perceived among practitioners that product form—and design in general (as the discipline responsible for the creation of new product form)—has become increasingly more important in the development of new products and services (Brown 2009, Maeda 2015). The jury’s initial nearly $1 billion (U.S. dollars) award to Apple in 2012 for Samsung’s infringement on design patents is testimony to this claim (The Economist 2012). And even though the literatures of marketing, engineering, and strategy have recently started to recognize the diverse roles that product form plays—for instance, in influencing perceptions of how customers value and understand a product (Bloch 2011), of how new product forms should be created and/or sourced (Erat and Krishnan 2012, Terwiesch and Loch 2004), of how the stock market can positively react to public evidence of a firm’s design effectiveness (Xia et al. 2016), and of how the links between technology and design should be managed (Rindova and Petkova 2007)—there is still no rigorous understanding of how product form evolves over time or how a new product form comes to predominate.

Businesses operate in a changing environment. Yet, with the possible exception of work addressing the fashion industry (e.g., Cappetta et al. 2006, Cillo and Verona 2008), we know of few studies that focus on how product form changes. Do changes in form follow changes in function, or not? Despite the century-old maxim of “form follows function” (Sullivan 1896, p. 408), extant research has not achieved conceptual closure on the matter (for various views, see, e.g., Eisenman 2013, Kreuzbauer and Malter 2005, Rindova and Petkova 2007), and neither is there large-scale direct empirical evidence to settle the question. Another leading question is, has form become more relevant in the creation of new products? If so,
then managers of new product development organizations need to adjust their organizational capabilities to changes not only in the technology landscape but also in the form factors of products that their firms develop (Rindova and Petkova 2007). However, the lack of large-scale quantitative data has prevented researchers from rigorously testing hypotheses about the role that product form might play in new product development.

Defining technology boundaries and product (or industry) categories has been instrumental in understanding how technology evolves (Utterback 1996). The equivalent notion with regard to product form is the style—a category of product designs that are similar in form. If we are able to categorize a body of individual designs into styles and develop an understanding of the temporal relations of these styles, then we can begin to address questions about the dynamics of product form. If we can also establish how styles are linked to functional categories, then we will have a solid empirical base that enables us to study the dynamics of product form and its relationship to changes in product function.

This paper offers two main contributions to the literature on technology and operations management. First, it makes available an unprecedented and rich data set of styles based on more than 350,000 new product forms that have been granted patent protection in the United States during the 33-year period starting in 1977. To identify these styles, we introduce a unique combination of state-of-the-art clustering techniques and experimental validation to categorize patents in the design patent database. In doing so, we have made the first step toward a large-scale categorization of styles by showing that they can be identified based on measures of pairwise similarity.

Second, having established this data set of styles, we advance the management literature by studying aspects of product evolution from the perspective of product form (rather than product function). In doing so, we focus on unexpected changes in styles—and hence of product form—because such changes have the potential to disrupt business activities and thereby attract the attention of managers and academics both. Using the notion of turbulence (Miller et al. 2006) to operationalize these unexpected changes, we study two key aspects of product form: we start by examining how turbulence in styles is related to turbulence in product functionality (using utility patents and functional product categories as a proxy for product function); we then study how style turbulence has changed over time. This approach reveals that turbulence in styles exhibits a U-shaped relation to turbulence in product functionality. The implication is that high levels of unpredictable changes in form are associated either with highly turbulent product functionality or with an utter lack of function turbulence. As a result, firms should treat turbulence in product form as a distinct source of uncertainty, because disruptions in product form can arise even in the absence of disruptions in product function. We also find that changes in styles have become increasingly unpredictable over time, which suggests that firms should rethink their organizational setups so as to cope with the increasing uncertainty. Such an approach would involve detecting changes in new product form, adopting development cycles that accommodate such changes, and setting up nimble production systems that can react to emerging form trends (Bourgeois and Eisenhardt 1988, Eisenhardt and Tabrizi 1995, Tece et al. 1997).

The rest of this paper proceeds as follows. Section 2 delves into how we operationalize “styles” for categorizing product designs. In Section 3 we discuss design patents and the unique features that make them ideal vehicles for an empirical study of product form. Sections 4 and 5 describe our methods—respectively, the graph-clustering approach employed to identify styles within the U.S. design patent database and the experimental approach used to validate our operationalization of style. In Section 6 we use the styles data set so constructed to study the dynamics of style turbulence and its relationship to turbulence in product function. We conclude in Section 7 by discussing the academic and managerial implications of our findings.

2. Operationalizing Styles in Product Design

Previous efforts to identify styles among a set of objects have aimed at explicitly identifying a style’s constituent physical design aspects. For instance, Munro (Munro 1946, p. 129) describes a style as consisting “of a combination of traits or characteristics which tend to recur together in different works of art.” He uses the example of Gothic architecture—which features pointed arches, pitched roofs, slender piers, large stained-glass windows, and flying buttresses—to illustrate the idea that a style is characterized by a set of well-defined physical traits. However, most researchers acknowledge that the mere co-occurrence of certain individual traits is less important than how those traits are configured or how they interact (Stacey 2006). In the field of engineering, efforts to capture and formalize not only a style’s defining elements but also the interactions among these elements have met with remarkable success. Consider shape grammar (Stiny 1980), a design language that consists of some basic geometric shapes together with a set of rules for transforming them into complex forms. This language has been used to describe the style of Harley-Davidson motorcycles (Pugliese and Cagan 2002) and of Buick automobiles (McCormack et al. 2004).

The marketing and engineering design literatures have developed approaches that study designs directly,
approaches that are tailored to specific product categories and hence implicitly take into account contextual information about those categories. For example, Cappetta et al. (2006) manually categorize a few thousand fashion designs into styles based on features such as cut, length, etc. derived from pictures of clothes in 228 issues of the magazine Vogue Italia. Jupp and Gero (2006) analyze styles within a set of 131 architectural designs by examining such characteristics as symmetry, regularity, and how shapes contain other shapes. Orth and Malkewitz (2008) group 160 wine bottle designs into styles along the lines of higher-order design elements, such as whether a design relates to nature. Landwehr et al. (2011, 2013) process 28 images of popular car models and examine how measures of prototypicality, complexity, and exposure predict sales. To identify styles in design patents, for which the corpus includes more than 350,000 designs across 33 product categories, we adopt a different approach that allows for massive scaling.

Since product form carries emotional and sociocultural meanings and since meanings are not measurable “objectively (without any human involvement)” (Krippendorff and Butter 1993, p. 33), it follows that operationalization in this case must ultimately rely on human judgment. Yet only human judgments that are relatively simple—and so can be deployed on a mass basis—can serve as the foundation of any broad effort to identify styles. We develop our approach by taking cues from research devoted to exploring how people perceive and understand complex objects. This research has developed the concept of cognitive categorization as one of its cornerstones (Simon 1969). In essence, humans simplify understanding of complex objects by categorizing them into groups; that process yields “archetypes” and thereby reduces cognitive load (Porac and Thomas 1990). Styles can be viewed as one such categorization.

This categorization process is, in turn, driven by the visual perception of similarities and differences (Holland et al. 1986, Rosch 1978). In the context of styles, such attributes are based on both individual features and their configuration (Stacey 2006); observers are intuitively able to make judgments of visual similarity on a holistic basis such that a single overall visual impression of similarity is reached (Goldstone 1994). There is empirical evidence that visual similarity can serve as a basis for recognizing styles (Chan 2000, Jupp and Gero 2006). Hence, we operationalize “styles” as categories of product form determined by perceived visual similarities.

Two comments are in order here. First, styles typically form a hierarchy. Munro (1946) proposes, in effect, a hierarchical structure when arguing that “restricted” styles (e.g., Florentine) are subsumed by “extensive” styles (e.g., Italian Renaissance). In this paper we focus on identifying the widest category of product form that humans would perceive to be a style. This category would be the root of a given style hierarchy; thus, it might subsume substyles but would not be subsumed by other styles. We refer to such a categorization as a main style. So by identifying main styles, we are able to study changes at the highest level of styles (Alexander 1964, Simon 1969).

Second, although individuals vary in their assessments of similarity, those who share a similar sociocultural background and basic level of expertise tend to produce comparable assessments that lead to agreed-upon categories (Goldstone 1994). Yet, because different sociocultural groups may apprehend designs differently, any categorization effort must identify the most appropriate reference group.

In sum, we operationalize styles as categories of visually similar product form that are organized in a hierarchical fashion. They are (i) established via a holistic perception of visual similarity such that each category contains designs that are more similar to each other than to designs in other categories and (ii) organized in a nested (hierarchical) fashion. Our methodological challenge is to group designs into design styles that enable the identification of what we refer to as main styles.

3. Design Patents
A novel design can be protected by filing a design patent with the U.S. Patent & Trademark Office (USPTO). Once the patent is granted, it protects the intellectual property related to the design’s visual characteristics—that is, the “appearance…which creates an impression through the eye upon the mind of the observer” (USPTO 2006, p. 1500-1). Design patents contain drawings that characterize the design; they also provide designer information, company information, location, date of filing, product category, and a list of references made to previous design works.

In contrast to utility patents, which protect the intellectual property concerning a product’s functional aspects, design patents protect its form aspects. To the extent that patents—especially the most frequently cited ones—are correlated with market success (see Hall et al. 2005), inventors and designers both have a commercial motive to seek patents for protection, especially since design patent rights improved significantly with the establishment of the Court of Appeals for the Federal Circuit in 1982 (Du Mont and Janis 2013). It is therefore not surprising that product designers are increasingly encouraged to “think about patents” when creating a new form (Molotch 2003, p. 28) and that many patent litigation cases have centered on designs; examples include Apple Inc. v. Samsung Electronics Co., 2011 WL 1523876 (N.D. Cal. Apr. 15, 2011) and Crocs Inc. v. Walgreens Co., 1:11-cv-02954-MSK (D. Colo. Nov. 14, 2011).
In Section 2 we noted that styles are socioculturally dependent and thus can be defined only with respect to an appropriate reference group. Although design patent law was not created with the intention of defining styles, it had to address the question of what was patentable and therefore needed to devise a criterion for determining whether (or not) a design was too similar to existing designs—that is, not sufficiently novel to merit patenting. Thus, the patent law had to identify who would assess “similarity” of visual appearance across product designs. In its landmark *Gorham Manufacturing Co. v. White* (81 U.S. 511) ruling of 1871, the U.S. Supreme Court established that design patentability should be determined by an “average observer” (rather than an expert) who possesses reasonable familiarity with the original and a (similar but) subsequent design were patented. The law’s reasoning for and choice of the average U.S. observer as the appropriate social reference group applies also to our task of identifying styles.

The patent examination process ensures that the “average observer” test is rigorously applied to every design patent application and that results are consistently documented in the patent. Every application undergoes examination to determine the design’s patentability. This process involves a patent examiner searching through a list of prior patents to find those that are similar in “visual impression” (USPTO 2006, p. 1500-29) to the applicant’s design. A patent application can be rejected if the resulting list contains a design that is substantially the same as the focal design. When an application is approved, the list of relevant patents found in the search process is documented in the patent documents as the list of references to prior works. This list of references constitutes the set of prior patents deemed most similar in visual concept to the focal patent.

The base properties of the design patents—namely, a singular focus on the visual character of designs, reference to the “average U.S. person” as the arbiter of visual similarity, and a rigorous process of identifying citations and hence similar designs—render such patents uniquely suitable for stylistic analyses. Furthermore, knowing the date when each patent was granted is an ideal setup for our examination of product form from an evolutionary perspective. Design patent data from January 1977 onward are available on the Internet courtesy of the USPTO (http://bulkdata.uspto.gov/). In our effort to identify styles, we examined all 357,305 design patents granted from January 1977 through January 2010.

### 4. Categorizing Designs into Styles

#### 4.1. An Index of Similarity in Form

A necessary step in categorizing designs into styles is to identify an index of similarity in form—that is, a measure of how close two designs are in terms of form. Because a patent application’s list of references is selected for visual similarity, such lists can serve as the basis for a measure.

However, there are three problems with directly using a list of references as a similarity index. First, the list is binary (either a reference exists or it does not), which means that it is a relatively coarse measure of similarity. Second, a new patent can cite an existing patent, but the existing patent cannot cite the new patent; thus, a measure based solely on references entails asymmetry, whereas similarity is a symmetric notion. Third, there may be incomplete relationships. For instance, it is typical for an entire year to elapse between the application for and the granting of a design patent. Another patent that is either granted or applied for during this period is less likely to be cited than are patents granted before the time of application (Chen et al. 2011). Our approach to constructing a similarity index from reference lists in patents tackles these issues via a heuristic for creating a single measure that is fine-grained, symmetric, and complete.\(^2\)

To devise a measure that is more fine-grained, we start by observing that a focal design whose patent cites many other patents draws its design inspiration from a large pool of extant work; such sourcing suggests that the focal design is relatively less similar (to any cited design) than in the case where only a few other patents are cited. Therefore, the first step of our heuristic is to (inversely) weight each individual reference listed in a patent by the total number of references in that patent’s list. Our second step in this heuristic is to impose symmetry by removing directionality from the references. These first two steps together yield a measure that we call *citation coupling*. The third step ensures completeness. For instance, similar patents filed at about the same time refer to (and are tested for patentability against) the same set of prior patents. Hence, we can strengthen our similarity measure by also determining the extent of overlap in the sets of references. Given two patents and their respective sets of references, we count the number of references that are common to both patents and then divide that number by the total number of references (without double-counting); the resulting quotient is a measure of the proportion of overlap in the sets. This method is known as *bibliographic coupling* (Kessler 1963). Finally, we sum the measures from citation coupling and bibliographic coupling to obtain a similarity index that ranges between 0 and 1.

This index is a score of the similarity in form between two designs. We can represent the entire set of relations...
as a similarity matrix. Alternatively, we can view this set as a weighted (nondirected) graph that spans all the designs (nodes of the graph) and where each edge of the graph constitutes a measure of similarity in form.

4.2. Clustering Method

Clustering naturally lends itself to the task of identifying styles in our similarity graph. The choice of a specific method depends on the data and assumptions made about cluster structures. Hence, we chose a conductance-based iterative divisive algorithm. Two features of our similarity graph are relevant to this choice: (i) the similarity graph exhibits more clustering than a random graph, and (ii) the distribution of node degrees (i.e., the sum of each patent’s similarity links) is highly skewed. From a theoretical standpoint, a hierarchically nested clustering structure would simultaneously exhibit both high clustering and skewed node degrees (Ravasz and Barabási 2003)—in line with our theory of styles as hierarchically organized categories. With a few technical assumptions, it can be shown that the iterative divisive algorithm proposed here can optimally recover such a hierarchical style structure (see Part A of the e-companion for a discussion and Part B for details of the algorithm).

Our algorithm takes the similarity matrix described in Section 4.1 as input (initially treating all 357,305 designs as belonging to one cluster) and performs clustering by iteratively partitioning the cluster into subclusters. The outcome is a hierarchical partitioning tree. Each iteration consists of

1. selecting a cluster to partition,
2. partitioning the selected cluster into two subclusters, and
3. evaluating whether the resulting clusters so formed constitute styles.

Intuitively, the ideal algorithm should select a heterogeneous cluster to partition at each stage; failing to do so would lead to unbalanced clusters, some of which would be heterogeneous and others homogeneous. It would also partition such that the generated subclusters exhibit both convergence (homogeneity within clusters) and divergence (heterogeneity across clusters); failing to do so would result in clusters having mixed content. Finally, the algorithm would ideally evaluate whether the clusters formed can be considered styles.

To operationalize heterogeneity, convergence, and divergence, our implementation leans heavily on the graph concept of conductance. Conductance unifies these properties into a single measure by explicitly conceptualizing a graph’s “heterogeneity” as the presence of a partition that can separate the graph into two parts such that each part is internally convergent but externally divergent (Chung 1997). Kannan et al. (2004) showed that iterative conductance-based algorithms identify clusters correctly if the data indeed contain clusters. Critically, they also show that the errors therein do not scale with the size of the data set—an important consideration in view of the large size of our data set. We use the so-called NJW algorithm of Ng et al. (2002) to calculate the conductance of clusters and to generate partitions.

In particular, we use conductance to operationalize the three steps of our algorithm. At the initial select step, we select the subgraph with the lowest conductance (i.e., the most heterogeneous subgraph) for partitioning; then, at the partition step, we use the partition implied by conductance. Finally, at the evaluate step, we identify sharp jumps in conductance between iterations; such jumps indicate a regime change in the underlying data structure, and the resultant grouping is therefore a prime candidate for the identifier of a main style. We identified five candidate solutions, corresponding to iterations 3,129, 5,749, 9,690, 15,463, and 22,065 (these iterations are labeled $O_1, \ldots, O_5$); see Part B of the e-companion for details.

5. Experimental Validation

In Section 2 we established styles as categories of designs based on a holistic visual perception of similarity. Hence, experiments with human subjects are well suited to validating the outcomes of our clustering approach. It is useful at this juncture to stipulate the three key assumptions that underlie our clustering approach, since the goal of our validation scheme is to test them.

i. Selection step: Conductance accurately measures how people perceive style heterogeneity in a set of designs.

ii. Partition step: The partition implied by conductance agrees with how people would categorize a set of designs into two groups based on perceived similarities and differences.

iii. Evaluation step: There exists some value of conductance beyond which a cluster of designs is recognized by people as a style (and this cutoff point identifies the main styles).

We note that if assumptions i and ii hold—that is, if the algorithm selects and partitions properly—then we can be assured that the hierarchical partitioning tree is properly identified. And if i and ii hold, then we also can test the claim of assumption iii that there exists an iteration (a level of conductance) beyond which clusters are recognized as styles. If this is true, then that iteration is the one that identifies the main styles.

To validate the selection step, we start by replicating—with human subjects—the selection tasks faced by the algorithm; then we compare the human and the algorithmic outputs. We perform this comparison in two ways. First, we test for whether there are nonrandom levels of agreement in the solutions that humans and the algorithm tend to propose. Second, we ask independent observers to assess the extent
to which the algorithm’s outcomes are different from those obtained from humans. Validation of the partition step proceeds in the same manner.

Finally, we validate that the clusters contained in one of the five candidate solutions are recognized by humans as styles. If at least one candidate solution is so recognized, then we expect subsequent candidate solutions (which contain still more homogeneous clusters) will also be recognized as styles. The first categorization that passes this test would be the one that establishes the main styles in our data set.

In short, our approach uses humans to validate not only the algorithm’s assumptions but also its outcomes. Validating that the algorithm performs its elemental tasks properly ensures its integrity, which is crucial if it is to be used in future studies.

5.1. Subject Pool

The subject pool for all of our experiments was recruited from Amazon’s Mechanical Turk (MT). The population of U.S. MT workers differs from the general U.S. population in that the former includes many more females (65%) than males (35%). The MT population is also younger (median age of 36), has a higher level of education, and has a lower income level than does the overall U.S. population (Paolacci et al. 2010). Despite these differences, the MT population still better represents the U.S. population at large than do student participants in a university lab setup (Paolacci et al. 2010). Also, as the next section makes clear, the experiments require a fairly large number of responses; MT provides a technical advantage over other approaches in the sheer number of respondents that can be gathered in a cost-feasible manner.

We restrict our sample in three ways. First, we limit the subject pool by requiring respondents to have a U.S.-based IP (Internet protocol) address—to fulfill the requirement that styles be perceived by an average U.S. person. Second, to ensure that MT subjects were attentive and properly understood the survey instructions, we implemented attention and comprehension checks in all surveys (see Part C in the e-companion for details) and then excluded respondents who failed any of the checks. (Our results, however, remain robust to their inclusion.) Finally, we do not reuse subjects across experiments.

5.2. Comparing Human and Algorithm Outcomes for the Selection Task

5.2.1. Human Replication of Algorithm Decisions. To compare the outcomes of the algorithm with those produced by humans, we have the latter perform the same selection tasks faced by the former. Specifically, we first sampled 25 selection tasks faced by the algorithm; because it performs exactly one selection task at each iteration (i.e., selecting the most heterogeneous cluster from a set of clusters), sampling a task is equivalent to sampling an iteration. We performed stratified sampling—in particular, we sampled five iterations from the iterations leading up to $O_1$ (the first candidate solution), five iterations from $O_1$ leading up to $O_2$, and so on up to $O_5$.

The algorithm considers numerous (in some iterations, thousands of) clusters when choosing one to partition during the selection phase. For that reason, humans cannot directly replicate a specific selection within a given iteration. To reduce the cognitive load, we simplify the selection task by asking each subject to pick only one out of three clusters. To find out whether the subject would (or would not) make the same choice as the algorithm, we sampled the cluster chosen by the algorithm together with two other random clusters that the algorithm did not choose; from the resulting three clusters, our subjects were tasked with identifying the most heterogeneous one (see Figure 1 for a sample question presented to subjects). Each subject performed this task for five samples. Altogether, we collected 185 valid responses from 37 respondents.

5.2.2. Test of Agreement. Given the algorithm and human outcomes, we can ask whether the algorithm and humans tend to converge to the same solution. To perform this analysis, we tested for whether or not the level of coincidence in solutions (i.e., the same solution being chosen by the algorithm and by humans) is better than chance. Note that since the selection involves three clusters, the probability of coincidence is equal to 0.33 if either the human or the algorithm does nothing more than pick clusters randomly.

We estimate the empirical probability of coincidence with the random model via a logit regression; the dependent variable is a binary indicator denoting coincidence, and we are interested in the probability of coincidence as captured by the size of the constant term. Our estimate of the empirical probability of a match is 0.48 (0.40–0.56; confidence interval estimated with errors clustered by respondent), which is significantly greater than random. The implication is that neither the humans nor the algorithm are performing randomly; because both have a better-than-random chance of choosing the same solution.

5.2.3. Turing Test. The agreement test alone is insufficient to validate the selection step: on those occasions where it does not match with human responses, the algorithm may be generating outcomes that a human would consider to be unreasonable. Intuitively, we need to test whether the algorithm is “human-like” with respect to the selection task—that is, does it ever exhibit significantly nonhuman behavior? Yet, because our tasks produce complex and unordered categorization outcomes, it is difficult to detect non-human-like
behavior. We address this challenge by deploying the Turing test (Turing 1950), where human judges attempt to detect nonhuman aspects of algorithm outcomes. The Turing test is meant to test the output of systems that are designed to mimic humans (Armstrong 2001, Barlas 1996, Sargent 1999, Van Horn 1971).

To implement the Turing test, we first supplement the 25 algorithm-generated solutions (where a “solution” is the outcome of one selection task—that is, three clusters with one identified as the most heterogeneous) used in the previous coincidence test with 25 human-generated solutions (randomly picked from the human responses in the previous test) to form a total of 50 solutions. We draw one solution at random from this group of 50 and show it to a human subject (i.e., one not involved in replicating the algorithm’s tasks). We task the subject with assessing whether the solution is of algorithmic or human origin (see Figure 2). If the subject can correctly identify outcomes generated by the algorithm, then those are clearly different from the outcomes generated by humans—here, the algorithm is not human-like in the sense that it exhibits significant nonhuman behavior, and so the validation fails. However, if the subject is unable to distinguish between algorithm and human outcomes, then we can reasonably presume that the former can generate human-like outcomes with respect to the selection task.

5.2.4. Statistical Inference Setup. Because the algorithm passes the Turing test if it is indistinguishable from humans, the Turing test is effectively a “null effect” test (Oppy and Dow 2016). Such tests rely on (i) a measure of the strength of the signal (of non-human-like behavior) exhibited by the algorithm and (ii) a statistical means of testing whether this signal strength allows for distinguishing algorithm and human.

Given that the algorithm should be indistinguishable from humans, many authors have found the probability that a judge correctly identifies an outcome’s source to be an intuitive measure of the signal strength (see, e.g., Church and Guilhardi 2005, Geman et al. 2015). More specifically, suppose that a judge is given the task of deciding whether a solution is generated by an algorithm (response \( r = A \)) or by a human (\( r = H \)). Correct answers occur when the response matches the true origin of the solution—that is, the judge should respond \( A \) to algorithm-generated solutions and should respond \( H \) to human-generated ones. Thus, we can write

\[
P(\text{Correct}) = P(A)P(r = A | A) + P(H)P(r = H | H);
\]

here, \( P(A) \) (respectively, \( P(H) \)) is the probability of the judge being presented with an algorithm (respectively, human) outcome. Note that our research design implies \( P(A) = P(H) = 0.5 \).

The intuition is that if humans are unable to distinguish between the two outcome types, then their responses should be random with respect to the underlying truth; in that case, \( P(r = A | A) = P(r = A | H) = p \), where \( p \) is an individual’s propensity to designate a presented outcome as algorithm-generated. Then

\[
P(\text{Correct}) = 0.5p + 0.5(1 - p) = 0.5.
\]

Yet if these two outcome types are easily distinguished, then \( P(\text{Correct}) \) should deviate from 0.5. So in this case, \( P(\text{Correct}) \) could either approach 1 (i.e., when the judge always identifies a solution’s origin correctly) or approach

**Figure 1.** Sample Question for the Selection Task

Identify the group that is the most heterogeneous with respect to styles. Note that a heterogeneous group would contain figures from different styles.
0 (when the judge never does so). Thus, the more that \( P(\text{Correct}) \) deviates from 0.5, the more it indicates a strong algorithmic signal—which means that we should reject the claim that the algorithm’s selection task performance is human-like.

Our testing strategy relies on previous work in the fields of economics, psychology, and medicine that aims to make statistically reliable statements concerning whether or not treatment effects are negligible (Cohen 1988, Solon 1992, Ziliak and McCloskey 2004). This work acknowledges that classical statistics cannot accept a null hypothesis. However, the essence of “null effect” testing (as is typical for the Turing test) is to show that any difference between humans and the algorithm is, at most, negligibly small; in other words, the algorithm is human-like in the sense that it generates at most negligible nonhuman signals (Shieber 2007). With respect to our selection task, this approach mandates that we first identify an interval around \( P(\text{Correct}) = 0.5 \) that we consider to be—for all practical purposes—“indistinguishable” from 0.5. In accordance with the most stringent standards for effect sizes in the medical and psychological literature (see, e.g., Cohen 1988, Ferguson 2009, Nakagawa and Cuthill 2007, Sawyer and Ball 1981), we require that \( P(\text{Correct}) \) fall within the range \( 0.5 \pm 0.1 \). Second, we must ensure that \( P(\text{Correct}) \) is within this range with at least 95% probability. The implication is that with at most 5% probability we make the error of claiming that an algorithm outcome is no different from a human outcome when, in fact, it is different. In short, the entire 95% confidence interval of \( P(\text{Correct}) \) must lie in the range \( 0.5 \pm 0.1 \).

When estimating \( P(\text{Correct}) \), we use the same logit specification as for the test of agreement in the previous section. To calculate the number of samples needed for an estimate of this precision, we performed a statistical power analysis (Cohen 1988) and found that guaranteeing the necessary confidence interval requires \( n = 384 \) responses. Formally, \( n = (2\sigma z_{0.975}/0.1)^2 \), where \( z_{0.975} = 1.96 \) is the \( z \)-score corresponding to a 5% two-tailed probability and where \( \sigma \) is the theoretical standard deviation of a Bernoulli response. To produce the most conservative (i.e., the largest) estimate of \( n \), we assume that \( \sigma = 0.5 \)—the largest value possible with a binary outcome distribution.

### 5.2.5. Results

We employed 404 respondents for the selection task, and each respondent performed 10 Turing tests. Because the first few answers for each respondent may be noisier owing to unfamiliarity with the task, in the statistical analysis we present results using only the last answer given by each respondent. We remark that our results are robust to picking any one of the 10 answers and also to including all 10 answers from each respondent (using respondent fixed effects). The value of \( P(\text{Correct}) \) is estimated as 0.46 (0.41–0.51), and the 95% confidence intervals are estimated with robust clustering on samples. Since \( P(\text{Correct}) \) is observed to be bounded within the range \( 0.5 \pm 0.1 \), it follows that our subject judges cannot accurately distinguish between algorithmic and human outcomes. Hence, human–algorithm differences in performing the selection task are negligible, and in this sense we claim that the algorithm is human-like (Shieber 2007).
5.3. Comparing Human and Algorithm Outcomes for the Partition Task

5.3.1. Human Replication of Algorithm Decisions. Just as we did for the selection task, here we must replicate the partition task using human subjects in order to build a basis for our statistical comparisons. For the partition task, the sampling procedure and the number of samples mirror those used for the selection task; thus, we sampled 25 partitioning tasks faced by the algorithm, 5 tasks for each of the five potential candidate solutions ($O_1$ to $O_5$). Because clusters are far too large for subjects to replicate the actual partitioning, this task is simplified by having them split 10 randomly chosen designs from a randomly chosen cluster into two subclusters of 5 designs each. Figure 3 illustrates the exact task. We had 40 respondents perform five partitions each to establish a pool of 200 human responses.

5.3.2. Test of Agreement. As before, our results are based on verifying whether the algorithm and the humans tend to converge to the same solution (again, by testing for better-than-random agreement). For the partition task there are 125 distinct ways of categorizing objects into two groups of five, which means that the probability of coincidence is only 0.008; our empirical estimate for coincidence yields 0.20 (0.14–0.25). These results imply that the partition task carried out by the algorithm coincides with results carried out by humans.

5.3.3. Turing Test. For the Turing test, we garnered 415 valid respondents (more than the $n = 384$ required by the power analysis). We obtain an estimate for $P(\text{Correct})$ of 0.53 (0.49–0.58). As was the case for our selection task, the algorithm does not produce unnatural results (i.e., those detectable by humans to be of algorithmic origin).

5.4. Identifying the Main Styles

So far we have established—for the selection and partition tasks, both—that the algorithm produces outcomes that overlap with humans and also that algorithm–human differences are, at most, negligible. Yet there is one task still to be completed. The algorithm selects and partitions iteratively: starting with the entire data set as one big cluster, the algorithm could in principle continue iterating until each design forms a unique, single-patent style. So where does the algorithm start producing clusters homogeneous enough to be considered styles? Because styles are hierarchically organized, multiple categorizations can be viewed as “containing styles.” However, our interest is in finding the main styles—which is equivalent to the first categorization (among the potential candidate solutions) under which the clusters match well with a human understanding of styles. This decision procedure can be operationalized via an easily interpreted criterion of simple majority: a candidate solution “matches well” when more than half of the population agrees that its clusters constitute styles.
Figure 4. Human Perception vs. Algorithmic Determination of Styles

We show below a group of 10 product designs. Please indicate whether you agree that the designs belong to the same style.

○ Agree  ○ Disagree

5.4.1. Experimental Design. For each of the five candidate categorizations, we sample 10 of the most heterogeneous clusters (as measured by conductance) to yield a total of 50 clusters. Each cluster is represented by 10 randomly sampled designs. We presented 10 randomly sampled clusters to a human subject and asked whether the designs in that cluster should be viewed as a style; the instructions and a sample are given in Figure 4. We obtained valid results from 233 respondents. To ensure independence of observations, we based the subsequent analysis on the subject’s answer to only one question; for this we used the last question, because that answer is probably less noisy. Our results are robust to basing the analyses on any one of the 10 answers as well as to including all 10 answers (using respondent fixed effects).

5.4.2. Results. To test whether respondents agree that the clusters representing each of the algorithm’s candidate solutions can be considered a style, we specify a logit model in which the independent variables are indicators for each of the candidate solutions. Figure 5 plots the marginal probabilities from that model. Observe that the first candidate solution that earns a simple majority vote is $O_3$. Points prior to $O_3$ (i.e., $O_1$ and $O_2$) contain clusters too broad to be viewed as styles; points after $O_3$ (i.e., $O_4$ and $O_5$) do contain styles, but, as explained previously, we are seeking the first categorization that passes this test. Hence, these results, which are based on a majority criterion, suggest that we use $O_3$ for identifying the main styles.

A second criterion—based on “structural breaks”—also supports the idea that clusters in $O_3$ constitute our main styles. Research in decision and psychology (e.g., Rosch and Mervis 1975) has identified structural breaks, or “regime changes,” as viable indicators of different categorizations. Figure 5 shows that there is a sharp and statistically significant structural break between $O_2$ and $O_3$; by contrast, $O_2$ and $O_1$, as well as $O_3$, $O_4$, and $O_5$, are statistically indistinguishable. Furthermore, and in line with our conceptualization of styles as a hierarchy of substyles, subjects recognize $O_4$ and $O_5$ (which further partition $O_3$) as styles. The fact that the level of agreement does not rise for finer partitions (the actual data display some noisy vacillations) reinforces the notion of a structural break (Mervis and Rosch 1981).

5.5. Summary of the Validation Steps

In sum, we have shown that the algorithm selects and partitions in a human-like fashion. Starting from a certain cutoff point, categorizations are recognized by humans as styles; this means that the cutoff point itself is a reasonable threshold for indicating main styles.

We can also inspect the algorithm’s clustering paths visually. The example presented in Figure 6 reproduces...
actual output derived from the algorithm and offers a visual summary of the validation tests conducted. In particular, we have shown that an average U.S. observer would

1. judge cluster \( O_3A \) to be more heterogeneous than cluster \( O_3B \),
2. partition cluster \( O_3A \) into cluster \( O_3A_1 \) and cluster \( O_3A_2 \), and
3. view cluster \( O_3A \) and cluster \( O_3B \) as main styles.

The outcome of our algorithm is a total of 9,690 styles covering more than 350,000 design patents. Figure 7 plots the size distribution of the styles. The graph indicates that design patents are distributed across both large “encompassing” and smaller “niche” styles. Thus our data set captures not only successful styles but also many failed styles that never reach the limelight.

The raw data support a second interesting conclusion. A comparison of our styles with the existing USPTO classification of designs reveals that they are separate constructs. The USPTO deploys a system that broadly classifies designs into 33 major product families. Although a style concentrates 80% of its designs (on average) in a single product family and hence has a dominant product family, a style typically spans 2.9 different product families. At the same time, each product family contains at least hundreds and often thousands of styles. There is also a finer USPTO classification of more than 5,000 subcategories such as “high chair for juvenile” and “simulative seating units” (USPTO 2005). Not surprisingly, this lower-level categorization further emphasizes the difference between our style construct and the USPTO classification—for which, on average, a style would contain designs coming from 8.4 different product subcategories, while a product subcategory would contain 15.5 different styles.

6. The Dynamics of Styles

Figure 8 plots the number of patents granted each year for the three most frequently occurring telephone handset styles—as identified from the data by our algorithm. The solid line represents designs of classic handsets, the dashed line represents designs of the candy-bar style (cell phones shaped like a block), and the dotted line represents designs of the clamshell style (cell phones that can be flipped open and closed). The graph shows the classic handset style enjoying a period of gradual growth leading up to a peak in patenting activity in 1986, followed by a long and slow decline. The candy-bar and (especially) the clamshell styles seem to follow a much more zig-zag pattern of growth. Examining how the prevalence of some established styles (such as those exhibited in Figure 8) evolves lends additional face validity to the outcome of our clustering approach. That is, the algorithm correctly accounts for the history of mobile phone design—identifying the best-known styles within the time frame considered—even though it is completely agnostic with respect to each patent’s time characteristics (since all temporal patterns are emergent) and all industry characteristics.

In addition to corroborating our algorithm with historical facts, Figure 8 suggests that year-to-year changes in the granting of design patents become more pronounced over time for each new style. Thus, changes in a style’s predominance might become less predictable. Turbulence captures this notion of unpredictable change (Miller et al. 2006). We define style turbulence as the amount of unpredictable changes—from year to year—in the popularity or “size” of a style (i.e., in the number of patents granted within that style).

It is intuitive that style turbulence measures the chance of today’s hot style suddenly going cold tomorrow or the reverse: a style with little activity suddenly gaining traction. In that sense, turbulence catches some firms off guard while creating opportunities for others. In general, high levels of turbulence demand flexibility from an organization’s internal structures, supply chain setup, and decision-making processes—placing greater emphasis on probing for change, willingness of management to switch tacks, and flexible design and production structures that can adapt to such changes (Eisenhardt 1989, Fine 1998).
Our data allow us to break new ground by linking style turbulence with two important questions. First, how is style turbulence related to turbulence in the underlying functionality of the patents associated with a given style? In other words, does style turbulence result mainly from the turbulence associated with the creation of new functionality, or is the relationship more nuanced? Second, are styles becoming more turbulent—and, hence, are design capabilities becoming more transient—over time?

6.1. Studying Style Turbulence

One might wonder why style turbulence would be related to function turbulence. One reason is that, even though styles reflect only the similarity of product form, those products can be viewed as a bundle of functions (Ulrich 2011). So, to the extent that styles are linked to product function, they are susceptible to changes in the functional domain. It is noteworthy that the relationship between function turbulence and style turbulence is nontrivial.

Because technological research and development tends to produce discontinuous changes in products, it also tends to precede and disrupt other aspects of products—for instance, industrial design (Vergyzer 2005). Beyond this “knock-on” effect from technological changes leading to directly visible form changes, even potentially invisible technological changes (such as the improved processing performance of a computer chip) may lead to changes to the product’s physical form, especially since form can be used to focus consumer attention on superior functional features (Hoegg and Alba 2011, Molotch 2003). This dynamic applies in particular to radically new technologies that depart from existing trajectories, because these technologies usually require that customers learn how to use them (Dougherty 1990, 2001) and such learning is facilitated by a well-designed product form (Hargadon and Douglas 2001). Hence, the introduction of radically new technologies often requires radically new forms, as both producers and customers seek to establish a common understanding of the new product (Rindova and Petkova 2007). This explains the results reported by Rubera and Droge (2013), who found that firms with high levels of both form and technological innovation do best. It follows that, if the designs of a given style are associated with technologies or functional domains experiencing high levels of turbulence, then the style itself is also likely to exhibit such turbulence. This argument is consistent with the widely held belief that “form follows function” (Sullivan 1896, p. 408).

That being said, styles can also be turbulent in the absence of changes in function, and a decidedly different mechanism may be at play for styles that are associated with stable functional regimes. At the extreme, the fashion industry (where technology seldom plays more than a marginal role) is a highly turbulent environment in which styles can change quickly and unpredictably from season to season—so much so that the entire high-fashion industry is organized around reducing the impact of such uncertainties (Godart 2012). Indeed, Kreuzbauer and Malter (2005) show experimentally that changes in form can influence how consumers perceive a product’s use and market category (e.g., off-road versus street motorbike)—implying that firms can rely on form changes (i.e., without major technological changes) to enable entry into related markets. Finally, technological stability may even facilitate experimentation in form by introducing technology platforms that reduce the cost of creating functionally equivalent products in various forms (Hölttä-Otto et al. 2008). So, when styles consist of designs that rest on stable functionality, the incentives to differentiate in form may become even higher and thus lead to highly turbulent styles.

These considerations suggest that the relationship between form and function is not linear (Eisenman 2013). In other words, turbulence in form can be triggered either by radical changes in function or by entrenched functional stability. We can express this idea formally as follows.

Hypothesis 1 (H1). A given style’s turbulence has a U-shaped relationship with respect to the function turbulence linked to that style.

Are styles themselves—that is, beyond the effect of function turbulence—likely to become more turbulent over time? Previous work has suggested that there is a fundamental rhythm or “clock speed” with which new products arrive to market (Fine 1998). For example, Mendelson and Pillai (1999) offered empirical evidence from the electronics industry that the rhythm of technological change is accelerating over time.

Two fundamental reasons may contribute to the increasing speed of new design introductions. First, the process by which firms deliver and market new designs is changing. Firms may leverage concurrent engineering (Loch and Terwiesch 1998), rapid prototyping (Thomke 1998), and open innovation (Terwiesch and Ulrich 2009) to become increasingly capable of delivering new designs in an ever-shorter time frame. Also, social media allows for instant feedback (Hildebrand et al. 2013) and can amplify the “contagion” effects of a viral design (Aral and Walker 2011). As a result, firms may become more capable of delivering new designs faster.

Second, there is evidence of a demand shift toward more innovative designs; in fact, such established companies as Sony, Apple, and Philips depend on bold designs for their success (Ravasi and Lojaco 2005). Such a shift may be the result of an increasing recognition of the importance of design (Maeda 2015). It may
also reflect a move away from technology as increasing levels of effort are required to produce the same extent of technological breakthrough (Jones 2009).

Taken together, these two effects could lead to a self-reinforcing cycle in which styles churn at an ever-faster rate. We therefore expect to observe an increase in style turbulence over time—even when function turbulence is taken into account. We thus posit our second hypothesis.

**Hypothesis 2 (H2).** Irrespective of function turbulence, style turbulence tends to increase over time.

### 6.2. Variables for Empirical Analysis

#### 6.2.1. Dependent Variable: Style Turbulence

The idea behind this measure is to separate those parts of style movements that are predictable from those that are unpredictable; the unpredictable parts are a measure for the uncertainty surrounding a style. To measure style turbulence, we follow Dess and Beard (1984) and define it as the “dispersion around a trend line, controlled for absolute size” (p. 58). In the spirit of that approach (as refined by McNamara et al. 2003), we calculate an index for a style’s turbulence as follows. First, we segment the style’s annual number of granted patents into four different five-year panels (1990–1994, 1995–1999, 2000–2004, and 2005–2009). Then, for each five-year panel of a style, we estimate a trend by regressing the number of patents granted yearly against a variable denoting the years in that panel. Finally, we obtain the dependent variable of interest as that regression’s standard error. Each observation yields \( T_{sp} \), or the turbulence of style \( s \) at period \( p \).

#### 6.2.2. Function Turbulence

To analyze the link between form and function, we need a predictor variable that reflects how each style is influenced by function turbulence—that is, the uncertainty regarding functions that are associated with a given style. Constructing such a predictor requires that we first identify functional domains with respect to which such turbulence can be defined and then link those functional domains to our styles. The USPTO defines domains of utility patents by grouping them into classes and subclasses based on proximate functionality (USPTO 2005, p. 3). Much as in Fleming and Sorenson (2004), we base our measure on the subclass of each utility patent (our results are also robust to using a class-based definition). Consistent with our measure of style turbulence, we conceptualize function turbulence as unexpected changes in the number of utility patents in these subclasses. Hence, we calculate function turbulence (for each utility category \( u \) in each period \( p \), denoted \( T_{up} \)) in the same way as for style turbulence. We link a specific style to a specific utility patent subclass by using citations from design patents to utility patents (design patents can cite utility patents, because form and function are not “easily separable” in practice; see USPTO 2006, p. 1500-02): when a style contains design patents that cite utility patents, the implication is that the form depicted in those design patents is influenced by the cited utility patents’ product functionality. Then, to calculate the total influence of function turbulence from the utility domain on a design style \( s \) in period \( p \), we identify the set of utility categories, \( U_{sp} \), to which any design patent in the style period refers (while allowing for repeated citations). We then calculate the total influence of function turbulence (on a style period) as \( \text{Function_turbulence}_{sp} = \sum_{u \in U_{sp}} T_{up} \).

#### 6.2.3. Style Activity

Given that style periods characterized by higher activity levels are likely to exhibit a higher standard error, we control for the mean number of patents granted yearly to the focal style by period. We use this variable to control for variations in activity levels across style periods.

#### 6.2.4. Product Categories

Finally, we remark that our data comprise products from multiple categories. There are 33 major product categories defined by the USPTO for the purpose of broadly grouping all design patents. Those categories may differ systematically in terms of market features (e.g., market size). As the market for a product increases, there could be more consumers willing to support a larger number of style changes. The patent document contains product category information (as for utility patents, those product categories defined on the basis of product function; USPTO 2005); hence, we control for market-related explanations by first identifying, for each style, the dominant product category (on average, a style has 80% of its designs from a single product category). We then add a time-varying control for the category activity (based on the total number of design patents in each product category in the period). We also control for any unobserved time-invariant characteristics of product categories by including fixed effects \( c_i \) for each product category \( i \). Table 1 reports the summary statistics and pairwise correlation of variables.

#### 6.2.5. Empirical Model

The full specification of our model is given by Equation (1). We make several observations as follows. First, all variables (except the time period) are logged because of their skewed distributions. Second, we use the single-period lagged variable \( \ln \text{Function_turbulence}_{sp-1} \). By using predetermined variables in the regression, we can establish whether function turbulence is predictive of style turbulence. Third, we include a quadratic term, \( (\ln \text{Function_turbulence}_{sp-1})^2 \), because we are positing a curvilinear relationship between the effect of function turbulence and the effect of style turbulence. The full specification is

\[
\ln T_{sp} = c_i + \beta_1 (\ln \text{Function_turbulence}_{sp-1}) + \beta_2 (\ln \text{Function_turbulence}_{sp-1})^2
\]
In Function_turbulence

Table 2. Summary Statistics of the Variables (N = 28,483 Style-Period Observations)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
<th>Mean (S.D.)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dependent variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. In Style_turbulence&lt;sub&gt;&lt;i&gt;p&lt;/i&gt;&lt;/sub&gt;</td>
<td>Log of the standard error of the linear regression for the annual style activity of style &lt;i&gt;s&lt;/i&gt; in period &lt;i&gt;p&lt;/i&gt;</td>
<td>0.72 (0.39)</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. In Function_turbulence&lt;sub&gt;&lt;i&gt;p-1&lt;/i&gt;&lt;/sub&gt;</td>
<td>Log of the sum of all turbulence of utility patent category receiving a citation from a design patent in a style &lt;i&gt;s&lt;/i&gt; in period &lt;i&gt;p - 1&lt;/i&gt;</td>
<td>1.66 (1.62)</td>
<td>0.46</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Period</td>
<td>Time period: 1 (1990–1994), ..., 4 (2005–2009)</td>
<td>2.51 (1.11)</td>
<td>0.16</td>
<td>0.34</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. In Style_activity&lt;sub&gt;&lt;i&gt;p&lt;/i&gt;&lt;/sub&gt;</td>
<td>Log of the mean annual patents granted in a style period</td>
<td>1.71 (0.96)</td>
<td>0.86</td>
<td>0.56</td>
<td>0.14</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>5. In Category_activity&lt;sub&gt;&lt;i&gt;p&lt;/i&gt;&lt;/sub&gt;</td>
<td>Log of the total activity of product category &lt;i&gt;i&lt;/i&gt; in period &lt;i&gt;p&lt;/i&gt; (to which the designs in style &lt;i&gt;s&lt;/i&gt; predominantly belong)</td>
<td>7.92 (0.82)</td>
<td>0.10</td>
<td>0.12</td>
<td>0.34</td>
<td>0.10</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Notes. All correlation coefficients are significant at the <i>p = 0.05</i> level.

\[
+ \beta_3(\text{Period}) + \beta_4(\text{In Style_activity}_{<i>p</i>}) + \beta_5(\text{In Category_activity}_{<i>p</i>}) + \varepsilon. \quad (1)
\]

We are mainly interested in the coefficients \(\beta_1, \beta_2,\) and \(\beta_3\); the first two specify the relationship between function turbulence and style turbulence, and the third specifies the time trend in style turbulence.

Note that our panels are wide (over 9,000 styles) but short (on average, each style is observed for only three five-year periods). Given these features, the appropriate estimation method is generalized least squares using robust errors clustered by styles. For wide panels, this method will yield asymptotically correct standard errors even in the presence of heteroskedasticity and serial correlation (Cameron and Trivedi 2009).

6.3. Analysis

Table 2 summarizes the regression results. We discuss the full model as specified in Equation (1). The coefficient for In Function_turbulence<sub><i>p-1</i></sub> is negative, whereas the coefficient for its quadratic term is positive; these results jointly suggest a convex relationship between form turbulence and function turbulence, which supports H1. To allow any arbitrary (stepwise) relationship between In Function_turbulence<sub><i>p-1</i></sub> and In Style_turbulence<sub><i>p</i></sub>, we also estimated a semiparametric model (Figure 9 plots the marginal effects of In Function_turbulence<sub><i>p-1</i></sub> and Table 2 reports the remaining coefficients); the graph confirms our hypothesized U-shaped relation.

Our full model also shows a positive and significant coefficient for Period. This finding indicates that, even after accounting for the effect of function turbulence, there is a steady increase in style turbulence over time. Thus, H2 is supported.

6.4. Robustness Tests

In this section, we present the results of five robustness tests performed under alternative specifications;

Table 2. Regression Model Predicting Style Turbulence (N = 28,483 Style-Period Observations)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Function only</th>
<th>Time only</th>
<th>Full model</th>
<th>Semiparametric</th>
</tr>
</thead>
<tbody>
<tr>
<td>In Function_turbulence&lt;sub&gt;&lt;i&gt;p-1&lt;/i&gt;&lt;/sub&gt;</td>
<td>–4.62*** (0.37)</td>
<td>–4.70*** (0.37)</td>
<td>(K_1, K_2, \ldots, K_s^p)</td>
<td>(K_1, K_2, \ldots, K_s^p)</td>
</tr>
<tr>
<td>(In Function_turbulence&lt;sub&gt;&lt;i&gt;p&lt;/i&gt;&lt;/sub&gt;)^2</td>
<td>0.81*** (0.09)</td>
<td>0.81*** (0.09)</td>
<td>(K_1, K_2, \ldots, K_s^p)</td>
<td>(K_1, K_2, \ldots, K_s^p)</td>
</tr>
<tr>
<td>Period</td>
<td>0.70*** (0.24)</td>
<td>1.30*** (0.24)</td>
<td>1.46*** (0.24)</td>
<td>1.46*** (0.24)</td>
</tr>
<tr>
<td>In Style_activity&lt;sub&gt;&lt;i&gt;p&lt;/i&gt;&lt;/sub&gt;</td>
<td>35.41*** (0.28)</td>
<td>34.92*** (0.32)</td>
<td>35.48*** (0.28)</td>
<td>35.58*** (0.29)</td>
</tr>
<tr>
<td>In Category_activity&lt;sub&gt;&lt;i&gt;p&lt;/i&gt;&lt;/sub&gt;</td>
<td>5.65*** (0.40)</td>
<td>2.21*** (0.89)</td>
<td>1.50*** (0.89)</td>
<td>1.33*** (0.89)</td>
</tr>
<tr>
<td>Product category fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>In Function_turbulence&lt;sub&gt;&lt;i&gt;p-1&lt;/i&gt;&lt;/sub&gt; dummies</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.75</td>
<td>0.75</td>
<td>0.75</td>
<td>0.76</td>
</tr>
<tr>
<td>F-statistic</td>
<td>631***</td>
<td>428***</td>
<td>617***</td>
<td>541***</td>
</tr>
</tbody>
</table>

Notes. Standard errors (in parentheses) are clustered by style. Coefficients and standard errors are scaled 100x.

\(^*^p < 0.05; ^*^*p < 0.01; ^*^*^p < 0.001.\)
see Table 3 for a summary. In all cases, we obtain significant support for our hypotheses.

First, our base model presumes that activity levels of the utility categories (i.e., the mean number of patents granted annually under the utility category $u$ in period $p$, denoted $M_{up}$), do not affect style turbulence. Yet our main regression has shown that turbulence and category size are highly correlated for individual styles. One might therefore argue that it would be preferable to use a predictor of function turbulence that is net of any activity effects. To allay such concerns, we control for the variable $\ln \text{Function}_u = \ln \sum_{p \in U_u} M_{up}$, which captures the influence of the total level of activity in a utility category, by summing $M_{up}$ over all utility patent categories that receive a citation from any design patent in the style (Model 1).

Second, we test a model where the controls for product categories are more refined (Model 2). We introduce dummies for each product category and for each time period; in essence, this allows also the unobserved product category–level characteristics to vary over time. In addition, we tested a model that allows the U-shaped curve to vary across product categories. We find a significant U shape in 12 of the 33 categories, and no category is significantly related to any other shape.

Third, we used a different method to calculate style turbulence (Model 3). Namely, given the annual number of patents of a style $s$ in period $p$ (denoted $\{x_{sp}^1, \ldots, x_{sp}^5\}$), we could alternatively calculate the turbulence of this style by assuming that the series evolves according to a random walk process with a linear drift; thus, $x_{sp}^n = x_{sp}^1 + u_n + \epsilon$. The unpredictability of this process is captured by the error term ($\epsilon$). We can estimate the standard error of $\epsilon$ by taking first differences and then calculating the sample standard deviation: $S_d(\epsilon) = \sqrt{\text{Var}(x_{sp})}$ (see, e.g., Cachon et al. 2007).

Fourth, our main model establishes that function turbulence is predictive of style turbulence one period later. We can establish a stronger form of predictive causality (Granger 1969) by including the lagged value of the dependent variable (i.e., $\ln \text{Style}_u$) in the regression (Model 4). This would help us determine whether function turbulence has predictive power beyond that implied by lagged values of style turbulence itself.

Fifth, we use instrumental variables (Model 5) to better establish the direction of our effects. We develop two instruments based on the idea of inventor churn. Arrivals of new inventors and departures of existing inventors clearly create disruptions in the technology world. However, if we assume that technology inventors do not patent designs and that designers do not patent technologies, then inventor arrivals and departures in the utility domain should not have any direct effect on style turbulence (and hence, inventor churn can affect only function turbulence, not style turbulence). Consequently, inventor churn can be an instrument to introduce exogenous change to style turbulence through function turbulence.

The patent documents contain the names of inventors, which are needed to create our instrumental

![Figure 9. Marginal Effects of Function Turbulence on Style Turbulence](image)

### Table 3. Results of Selected Robustness Tests ($N = 28,483$ Style-Period Observations)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\ln \text{Function}_u$</td>
<td>-10.13***</td>
<td>-4.51***</td>
<td>-2.81***</td>
<td>-5.32***</td>
<td>-4.93***</td>
</tr>
<tr>
<td>($\ln \text{Function}_u$)²</td>
<td>1.14***</td>
<td>0.77***</td>
<td>0.60***</td>
<td>0.83***</td>
<td>0.87***</td>
</tr>
<tr>
<td>Period</td>
<td>1.27***</td>
<td>Absorbed in FE</td>
<td>1.11***</td>
<td>1.34***</td>
<td>1.27***</td>
</tr>
<tr>
<td>$\ln \text{Style}_u$</td>
<td>35.52***</td>
<td>35.47***</td>
<td>44.31***</td>
<td>34.98***</td>
<td>35.40***</td>
</tr>
<tr>
<td>$\ln \text{Function}_u$</td>
<td>2.54***</td>
<td>Absorbed in FE</td>
<td>2.43*</td>
<td>1.50 (0.88)</td>
<td>1.50 (0.88)</td>
</tr>
<tr>
<td>$\ln \text{Category}_u$</td>
<td>1.26 (0.88)</td>
<td>Absorbed in FE</td>
<td>2.43*</td>
<td>1.50 (0.88)</td>
<td>1.50 (0.88)</td>
</tr>
<tr>
<td>$\ln \text{Style}_u$</td>
<td>3.39***</td>
<td>Absorbed in FE</td>
<td>1.50 (0.88)</td>
<td>1.50 (0.88)</td>
<td>1.50 (0.88)</td>
</tr>
<tr>
<td>Category fixed effects</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Category-period fixed effects</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.76</td>
<td>0.76</td>
<td>0.77</td>
<td>0.75</td>
<td>0.75</td>
</tr>
<tr>
<td>F-statistic</td>
<td>604***</td>
<td>182***</td>
<td>970***</td>
<td>677***</td>
<td>588***</td>
</tr>
</tbody>
</table>

Notes. Standard errors (in parentheses) are clustered by style. Coefficients and standard errors are scaled 100x.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. 

Downloaded from informs.org by [163.119.58.180] on 15 March 2018, at 09:46. For personal use only, all rights reserved.
variables. We rely also on the disambiguation of names into unique inventor IDs by Li et al. (2014). We start by creating two variables for each period $p$ of a utility category $u$. The first variable is $\text{Arrival}$, defined as the number of unique inventors observed in category $u$ for period $p$ but not for $p-1$. The second is $\text{Departure}$, defined as the number of unique inventors observed in category $u$ for period $p-1$ but not for $p$. We then implement a two-stage least-squares model with the two (logged) variables as instruments for $\ln \text{Function turbulence}$.11 We note that either of these instruments can be used in isolation. More crucially, neither instrument is weak: the $F$-statistics for these variables—7,889 for $\ln \text{Arrival}$ and 8,053 for $\ln \text{Departure}$—are much higher than the critical value of 10 (Stock et al. 2002). Finally, the Hansen test does not reject the null hypothesis that the instruments are exogenous: Hansen’s $J = 0.22 (p = 0.90)$.

Finally, we also tested (but do not report here) our results using different parameter choices—including 4-year and 10-year instead of 5-year time windows, main utility categories rather than subcategories, styles identified by $O_4$ instead of by $O_3$, and alternative lag structures (i.e., no lags or controlling for multiple lags of $p-1$ and $p-2$). Our results are robust to all of these alternative treatments.

7. Discussion

It is of great interest to understand how products evolve and progress. Business managers seek to understand past events and to predict future events so that they can steer their company appropriately. As a result, this topic has spawned a wide literature on technology evolution that has found its way into the toolkits of many business decision makers and consultants (Baldwin and Clark 2000, Utterback 1996). Yet, that literature has viewed progress in products almost exclusively from the technology standpoint—in other words, from the perspective of product function. Although functionality is arguably the most important factor in determining how products evolve, designers and marketers have long argued that products should be viewed as bundles of function and form (Bloch 1995, Ulrich 2011). Thus, product evolution might be determined not only by functionality but also by form.

Despite product form having recently garnered more attention in business circles (Ravasi and Lojacono 2005, Verganti 2006) and even though there is a large body of work in the field of marketing that focuses on how consumers perceive and value designs, the management of design has not received sufficient interest in the academic management community with regard to any topic other than marketing (Noble 2011). Although many reasons can be advanced to explain this deficiency, an important factor is the lack of an empirical basis for rigorous, large-scale studies focused on product form. The first contribution of our paper is thus to render product form (and design, as the discipline that leads the creation of new forms) amenable to empirical research by making available a broad data set on styles. We achieve this by (i) identifying styles in the USPTO design patent database through a rigorous conceptualization of “design style,” (ii) deploying a state-of-the-art clustering algorithm, and (iii) using a set of experiments to verify rigorously the algorithm’s output.

Our styles data set features three important properties that make it a useful platform for studying the role of product form in new product development. First, the styles data set disentangles product form from product function because it is built from design patents, which systematically capture a product’s novel form factors; hence, their citation patterns establish unambiguous evidence of form similarity among design patents. As a result, the styles presented here are new entities formed by visually similar products (i.e., irrespective of their functionality). Second, our styles data set is based on the hundreds of thousands of design patents granted in the United States during the period 1977–2010, which enables us to examine the dynamics of styles over three decades. Third, because design patents capture diverse information about their creation—such as the time of patent application, the link to the technology that underlies the product, and information about designers—the USPTO database provides a rich empirical basis on which to advance our understanding about the creation of product form (i.e., beyond the influence of product function).

A second contribution of this paper consists of using our styles data set to study what drives the unpredictability of changes in product form—that is, style turbulence. Style turbulence captures the inherent uncertainty associated with changes in product form, and understanding its drivers yields insights on managing the risks associated with the creation of a new product form. After examining the effect of function turbulence on style turbulence, we find that highly turbulent styles can be associated with products whose functionality is either turbulent or stable. This finding leads us to conclude that the relationship between form and function is nontrivial; in particular, it is certainly not always the case that form follows function (Veryzer 2005).

The U-shaped relation that we identify between function turbulence and form turbulence also speaks to the literature on a product’s life cycle (Utterback and Abernathy 1975). In particular, the S-curve view of product evolution postulates that dynamism might decline during the later phases of a product’s (or industry’s) life. Yet our findings indicate that product categories or industries can avoid becoming less dynamic by shifting their source of dynamism from function to
form, thus extending their life cycles. Indeed, in a further analysis of the trends of function turbulence and style turbulence over time, we find corroborating evidence that styles associated with relatively stable technologies tend to see decreasing function turbulence (reflecting technology maturing) and increasing style turbulence (reflecting shifts of dynamism to styles). The computer industry during recent decades is a vivid example of this assertion (Maeda 2015). Hence, investments in changing product form—and managing its uncertainty—could become new managerial imperatives during the waning phase of a product’s life cycle.

Our analysis also reveals that—irrespective of the functional domain’s influence—the extent of style turbulence is increasing over time. That is, we continue to see more aesthetic “churn.” Because this analysis accounts for the effect of changes in product function, our findings on the increasing time trend of style turbulence indicate that unexpected changes in product form occur at a faster rate than do changes in product function. The implication is that managing the uncertainty and risk associated with product form is more important now than ever before. Practitioners have long been offering anecdotal evidence of the increasing importance of design (Brown 2009, Maeda 2015). We show rigorously that this increasing importance for the development of products is not simply a matter of some companies paying more attention to this aspect; rather, it is a general trend across industries—more than half of the product categories in our data set exhibited increasing style turbulence, and none of the categories exhibited a significantly decreasing trend.

More broadly, the results of this research constitute evidence that product managers need to depart from the traditional view that product function is the main driver of product evolution and also to consider product form as an important source of both uncertainty and opportunity when developing new products. To the extent that the resources, competencies, and processes employed in product design activities differ from those employed in technological design activities, turbulence in product form can introduce significant challenges to managing these capabilities. This means that adaptation to increasingly turbulent styles requires more flexible design processes allowing for more rapid iterations of form; it may also require new product architectures to allow for greater design flexibility. For example, many automobile companies have recently employed “platforming” to separate a vehicle’s core mechanical and electronic components from its “hat” (i.e., the parts that are visible to a consumer). Any such changes need to be synchronized with the firm’s production systems, including supplier structure and supplier management. When product cycles indicate faster design churn, marketing cycles should be adjusted accordingly. Firms may ultimately need to elevate the presence of design in the organizational structure so as to reflect these new, design-driven realities.

There are, of course, some limitations to our method of identifying styles. First, although our experimental comparisons of algorithm-generated versus human-generated outcomes indicate that the human–algorithm differences are negligible, these comparisons are based on (random) samples of a size that humans can comfortably handle. Second, our approach to identifying styles does not rely on identifying the specific physical or psychological features of each style. Yet we believe that our work provides an important empirical basis upon which—in combination with other techniques (such as shape grammar, psychological research, conjoint analysis, etc.)—each style can be substantially interpreted. Third, we used patent data to disentangle the form and function aspects of products; more work would be required to link such data to actual products and thereby devise more direct measures of product success.

Despite its limitations, we envision this paper serving as a research platform. The styles data set that we have assembled opens opportunities for researchers to investigate, in a rigorous way and on a large scale, questions about product form. Do designers and design firms concentrate their efforts within certain styles? To what extent are star designers instrumental in the creation of new styles? Do design hubs, such as Southern California and New York, play a role in a style’s emergence? And are styles initiated there, or are those areas primarily marketing hubs? Which style adoption strategies most improve the firm’s financial viability? The styles data set described in this work offers a foundation on which the research community can build to develop an evidence-based perspective on the management of product form and design. We welcome all researchers to use the database and thus to participate in building such a perspective.

Acknowledgments

The authors thank department editor David Hsu, the associate editor, and three referees for their constructive feedback. They also thank Haibo Liu, Serguei Netessine, Phanish Puranam, and Jasjit Singh, all of whom provided valuable comments to earlier versions of the paper. The authors also appreciate the feedback from participants at seminars at Bocconi University, Emory University, Frankfurt School of Finance & Management, Georgetown University, Hong Kong University, Kühne Logistics University, London Business School, Massachusetts Institute of Technology, New York University, Singapore University of Technology and Design, the University of California at Los Angeles, the University of California at San Diego, and the University of Virginia.
Endnotes

1 Although the design patent examiner is likely to have much more experience than an average observer, patenting examination procedures provide clear guidelines for how the examination should proceed (USPTO 2006). The design patent examiner is trained in the capacity to observe designs as would an average person. Design patent examiners also follow a thorough search process: every design is listed in one (or more) product classes, and examiners typically search across multiple classes for related designs; moreover, designs may be separately annotated with search notes to make sure that the search is comprehensive (USPTO 2005).

2 One feature of the design patent is that designs in the corpus are unique—since designs that are similar enough to confuse the average observer are not considered patentable and so never appear together in the USPTO database. An investigator using a database that contains essentially identical designs should consider an additional pre-processing step to merge those designs, because designs that are essentially identical belong by definition to the same style and hence should not undergo clustering.

3 We draw inspiration from Kornish and Ulrich (2011), who faced a similar challenge of identifying clusters of unique ideas; they applied a clustering method after establishing a similarity measure across ideas.

4 An agreement probability of 0.20 is not low, given the strict requirement that all items must categorize exactly. If we relax the matching criterion and allow for slight deviation—for example, allowing one item to be miscategorized (as would be very common when we compare human answers)—then the probability of agreement increases to 0.60.

5 We replicated all the validation results while controlling for task difficulty (not presented here). Easier tasks have a weak ($p < 0.10$) effect on increasing levels of agreement but have no effect on the Turing test. All of our insights are robust to this alternative formulation.

6 A random sampling over all clusters would tell us whether the average cluster in a categorization is considered a style. However, we perform this test in a more robust manner by sampling over the most heterogeneous clusters in each categorization (i.e., those least likely to be judged as styles). Operationally, we define as “heterogeneous” those clusters identified for partitioning before the next candidate solution is reached (this amounts to about 20% of the most heterogeneous clusters in each candidate solution).

7 Because our data end at the beginning of 2010, “smartphones” has not yet become one of the top three design styles. Even so, after 2007, our clustering approach identifies smartphones as a growing style.

8 For the analysis undertaken here, we use data from 1985 through 2009, because design patent protection was weaker until 1982 (Du Mont and Janis 2013). We also lose one period as a result of our use of a lag structure in the main model; hence, we can use only four periods of five years each. For this we select the two most recent decades. Our results are robust to specifying 4-year or 10-year panels instead. However, extremely long panels (i.e., more than 10 years) may bias the measure upward because bona fide trends lasting fewer than 10 years may be wrongly interpreted as turbulence.

9 We have also tested a model that applies fixed effects to all the product categories to which any design in a style belongs. Our results are robust to this alternative formulation.

10 We drop observations where a style exhibits no activity, because inactivity perfectly predicts zero turbulence, in which case the other variables have no informational value.

11 Note first that the regression setup also includes a quadratic term for function turbulence, which by extension could also be endogenous; hence, for any instrument $X$ that we identify, we also include $X^2$ as an instrument (Wooldridge 2010). Second, function turbulence is the sum of turbulence over all relevant utility categories (indicated by citations from the focal style), so for the identified instruments, we similarly sum over the same categories. Finally, we use the same one-period lag for the instrument as we do for function turbulence.

References

Cameron AC, Trivedi PK (2009) Microeconometrics Using Stata (Stata Press, College Station, TX).


