INSEAD
The Business School for the World®

Working Paper Series
2016/40/TOM
(Revised version of 2015/49/TOM)

On Styles in Product Design: An Analysis of US Design Patents

Tian Chan
INSEAD, tian.chan@insead.edu

Jurgen Mihm
INSEAD, jurgen.mihm@insead.edu

Manuel Sosa
INSEAD, manuel.sosa@insead.edu

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 US 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.

Keywords: Product Form; Cluster Analysis; Experiments; Design Patents; Industry Turbulence

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 (US) 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), 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 and, in particular, 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), 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 as regards 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 their temporal relations, then we could begin to address questions about the dynamics of product form. If we could also establish how styles are linked to functional categories, then we would have a solid empirical base that would enable 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. In order 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 and Glick 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. That approach would involve detecting changes in new product form, adopting development cycles that accommodate such changes, and setting up a nimble production system that can react to emerging form trends (Bourgeois and Eisenhardt 1988, Eisenhardt and Tabrizi 1995, Teece 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 US 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 (1946) 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 defined by a set of well-defined physical traits. However, most researchers acknowledge that the mere co-occurrence of certain individual traits is no more 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 their interactions 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) consider styles based on individual features of fashion designs (cut, length, etc.) and manually categorized a few thousand designs—based on pictures of clothes in 228 issues of the Italian *Vogue* magazine—into styles. 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 proto-typicality, 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 category.

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 both on 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. That 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 US 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 an observer" (USPTO 2006, pp. 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 2011). It is therefore not surprising that product designers are increasingly encouraged to "think about patents" when creating a new form (Molotch 2004, p. 28) and that many patent litigation cases have centered on designs; examples include *Apple Inc. v. Samsung Electronics Co.* (2011) and *Crocs Inc. v. Walgreens Co.* (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 v. White* ruling of 1871, the US Supreme Court established that design patentability should be determined by an "average observer" (rather than an expert) who possesses reasonable familiarity with the designs (USPTO 2006, pp. 1500–21). The logic underlying this decision was its commercial impact: if a product's potential buyer confuses its design with a prior design, then the original design's inventor could suffer a commercial loss in the event that both the original and a (similar but) subsequent design were patented. The law's reasoning for and choice of the average US 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, pp. 1500–29) to the applicant's design.[1] 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

[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).

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 US 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 is available on the Internet courtesy of the USPTO. In our effort to identify styles, we examined all 365,444 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 their visual similarity, such lists can serve as the basis for a measure.

However, there are three problems with directly using the 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 (Lei and Wright 2009). Our approach to constructing a similarity index from reference lists in patents resolves these issues via a heuristic for creating a single measure that is *fine-grained*, *symmetric*, and *complete*.[2]

In order 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 single 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 heuristic's second step 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.

[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 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.

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.[3] 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 365,444 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.

[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.

In order 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 and also that their errors do not scale with the size of the data set—an important consideration in view of our data set's large size. We use the so-called NJW algorithm of Ng, Jordan, and Weiss (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 iteration 3,129, 5,749, 9,690, 15,463, and 22,065 (these iterations are labeled $O_1, \ldots, O_5$, respectively); 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.

In order 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 US MT workers differs from the general US population in that the former includes many more females (65%) than males (35%). The MT population is also younger (median age 36), has a higher level of education, and has a lower income level than does the overall US population (Paolacci et al. 2010). Despite these differences, the MT population still better represents the US 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 US-based IP (Internet protocol) address—to fulfill the requirement that styles be perceived by an average US person. Second, to ensure that MT subjects paid attention 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 results from 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

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 also 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. In order 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 if 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.

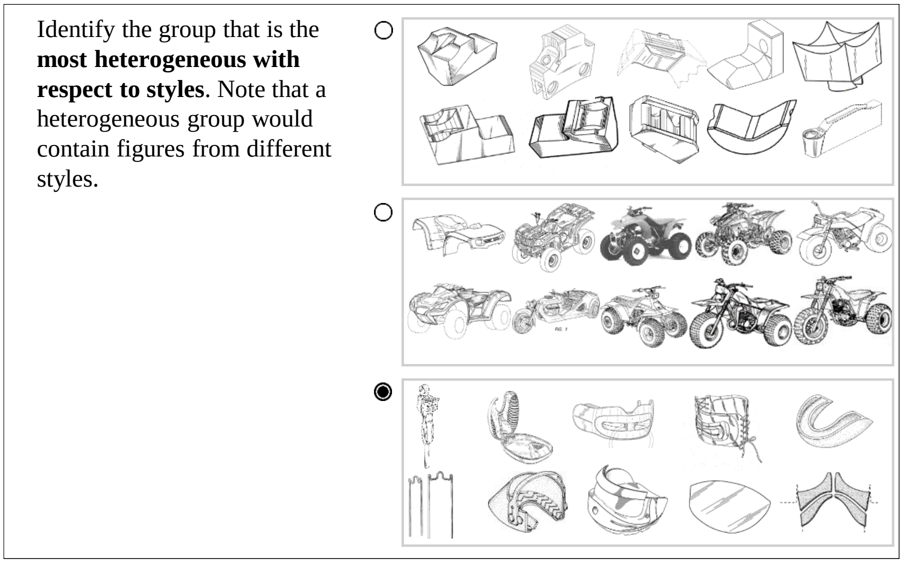

Figure 1: Sample question for the selection task

Test of agreement. Given the algorithm and human outcomes, we can ask if 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.

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 non-human 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 non-human aspects of algorithm outcomes.

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 non-human 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.

Statistical inference setup. The Turing test has long been used to test the output of systems that are designed to mimic humans (Armstrong 2001, Barlas 1996, Sargent 1999, Van Horn 1971). Because the algorithm passes if it is indistinguishable from humans, the Turing test is effectively a "null effect" test (Oppy and Dowe 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 algorithm-generated (response $r = A$) or human-generated ($r = H$). 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 $P(A)$ (resp. $P(H)$) is the probability of the judge being presented with an algorithm (resp. 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(Correct) should deviate from 0.5. So in this case, P(Correct) could either approach 1 (i.e., when the judge always identifies a solution's origin correctly) or approach 0 (when the judge never does so). Thus the more that P(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 non-human signals (Shieber 2007). With respect to our selection task, this approach mandates that first of all we 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*(Correct) fall within the range 0.5 ± 0.1. Second, we must ensure that *P*(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*(Correct) must lie in the range 0.5 ± 0.1.

When estimating *P*(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 σ 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.

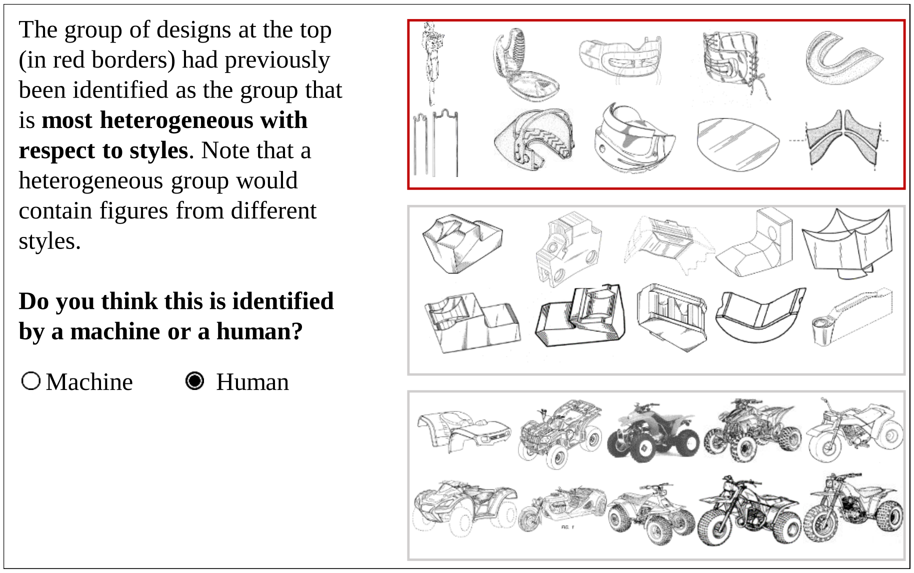

Figure 2: Sample question for the Turing test (selection task)

Results. We employed 404 respondents for the selection task, and each respondent performed ten 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 ten answers and also to including all ten answers from each respondent (using respondent fixed effects). The value of *P*(Correct) is estimated as 0.46 (0.41–0.51), and the 95% confidence intervals are estimated with robust clustering on samples. Since *P*(Correct) is observed to be bounded within the range 0.5 ± 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

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 mirrors those used for the selection task; thus we sampled 25 partitioning tasks faced by the algorithm, five 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 ten randomly chosen designs from a randomly chosen cluster into two subclusters of five designs each. Figure 3 illustrates the exact task. We had 40 respondents perform five partitions each to establish a pool of 200 human responses.

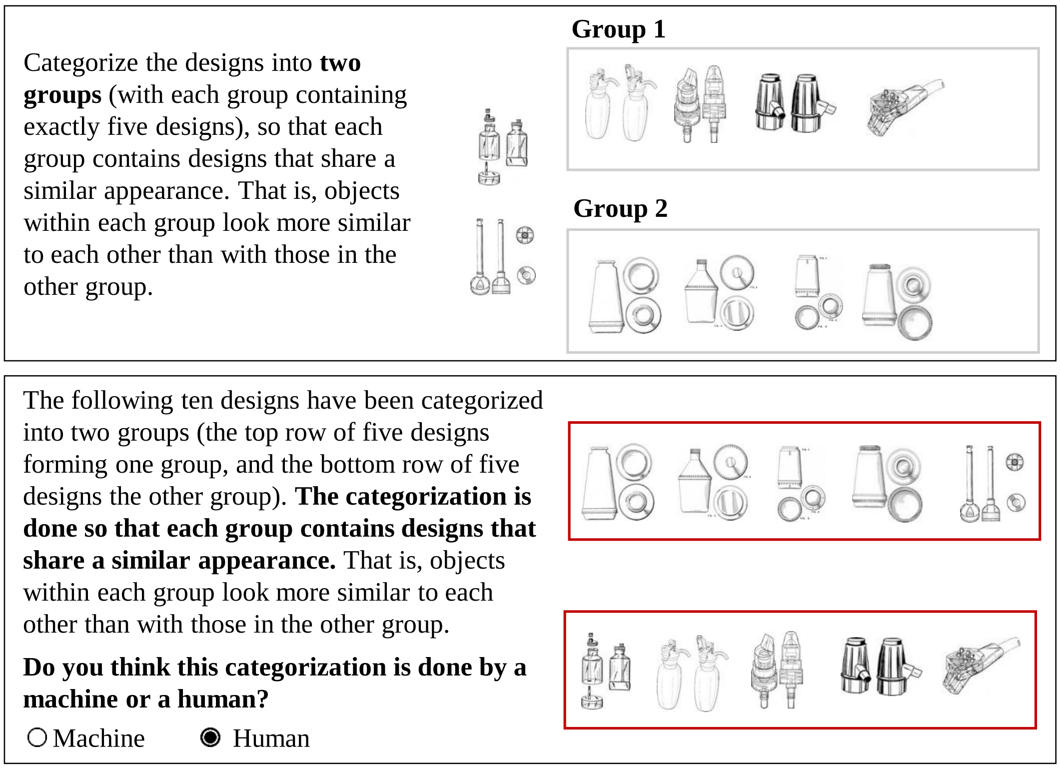

Figure 3: Sample questions for the partition task (top) and its Turing test (bottom)

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).[4] These results imply that the partition task carried out by the algorithm coincides with results carried out by humans.

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(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]

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.

Experimental design. For each of the five candidate categorizations, we sample ten of the most

[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 our insights are robust to this alternative formulation.

heterogeneous clusters (as measured by conductance) to yield a total of 50 clusters.[6] Each cluster is represented by ten randomly sampled designs. We presented ten 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 ten answers as well as to including all ten answers (using respondent fixed effects).

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

○ Agree ◉ Disagree

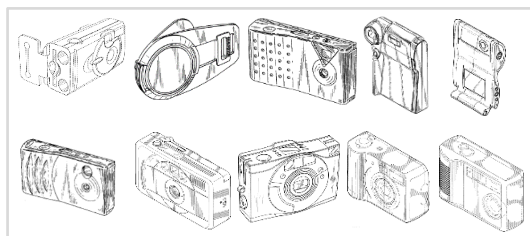

Figure 4: Human perception versus algorithmic determination of styles

Results. In order 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.

[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).

Hence these results, which are based on a majority criterion, suggest that we use O_3 for identifying the main styles.

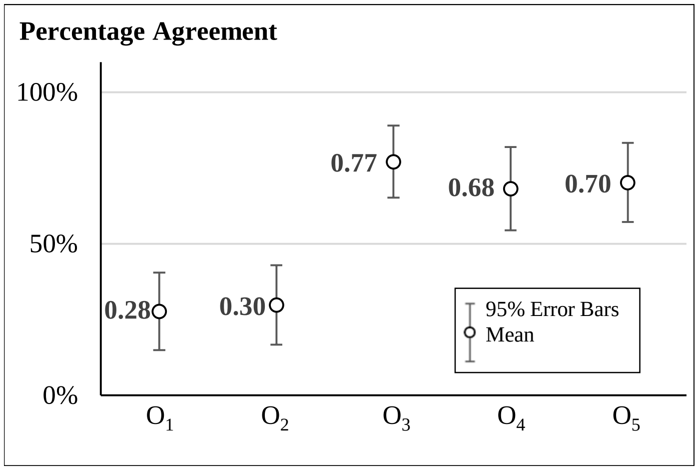

Figure 5: Percentage of subjects who agree that "these designs are from a single style"

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; in 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. 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.

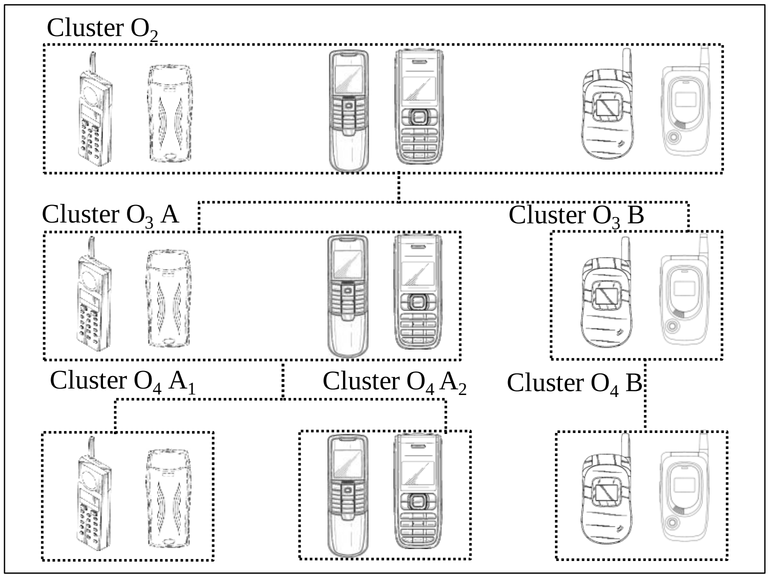

Figure 6: Example of a hierarchical partitioning tree generated by the algorithm

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 US observer would:

1. judge Cluster O_3A to be more heterogeneous than Cluster O_3B;
2. partition Cluster O_3A into Cluster O_4A_1 and Cluster O_4A_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 over 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.

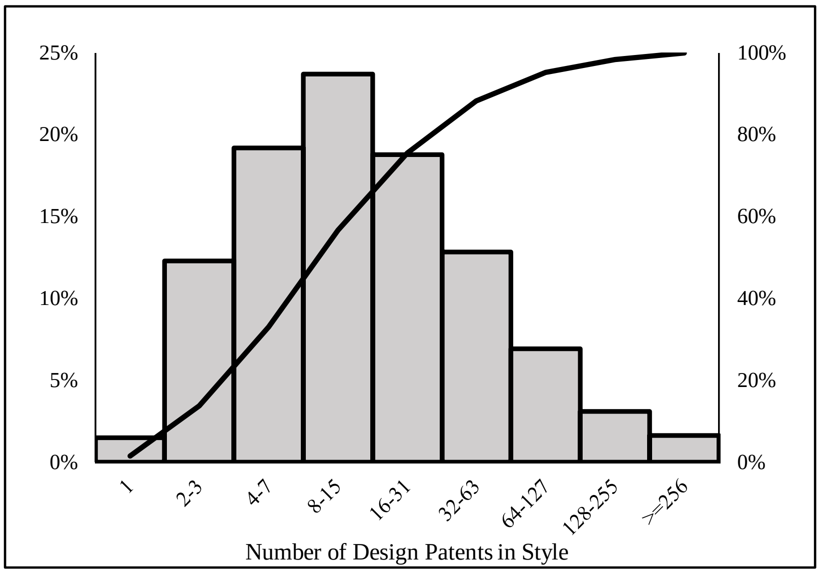

Figure 7: Size Distribution of Styles

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-zagged pattern of growth.

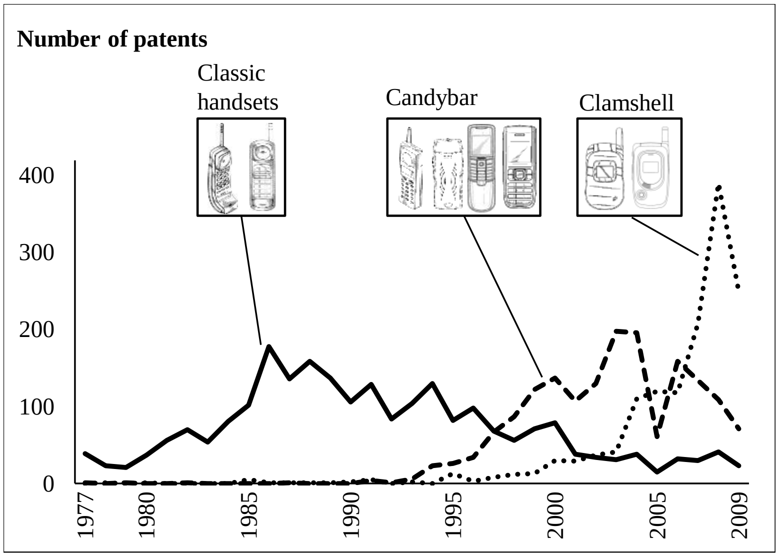

Figure 8: Number of patents granted annually for the three most popular telephone styles

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[7]—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 "jagged" 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 and Glick 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

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

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 (Veryzer 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 2004). 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 that radically new forms be tried 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).

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 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 Lojacono 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

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).[8] 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.

Function turbulence. In order 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

[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 2011). We also lose one period owing to 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 four-year or ten-year panels instead. However, extremely long (more than ten-year) panels may bias the measure upward because bona fide trends lasting fewer than ten years may be wrongly interpreted as turbulence.

"easily separable" in practice; USPTO 2006, pp. 1500–2): 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, in order to calculate the total influence of turbulence from the utility domain on a design style s in period p, we identify the set of utility categories, $\boldsymbol{U}_{sp}$, 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 $Function\_turbulence_{sp} = \sum_{u \in \boldsymbol{U}_{sp}} T_{up}$.

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

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 can 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.[9] Table 1 reports the summary statistics and pairwise correlation of variables.

[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.

Table 1: Summary statistics of the variables (*N* = 28,483 style-period observations

Variable	Description	Mean (S.D.)	Correlation Matrix 1	2	3	4	5
Dependent Variable							
1. ln $Style\_turbulence_{sp}$	Log of the standard error of the linear regression for the annual style activity of style *s* in period *p*	0.72 (0.39)	1.00				
Independent Variables							
2. ln $Function\_turbulence_{sp-1}$	Log of the sum of all turbulence of utility patent category receiving a citation from a design patent in a style *s* in period *p* – 1	1.66 (1.62)	0.46	1.00			
3. *Period*	Time period: 1 (1990-1994), …, 4 (2005-2009)	2.51 (1.11)	0.16	0.34	1.00		
4. ln $Style\_activity_{sp}$	Log of the mean annual patents granted in a style period	1.71 (0.96)	0.86	0.56	0.14	1.00	
5. ln $Category\_activity_{ip}$	Log of the total activity of product category *i* in period *p* (to which the designs in style *s* predominantly belong)	7.92 (0.82)	0.10	0.12	0.34	0.10	1.00

Note: All correlation coefficients are significant at the $p = 0.05$ level.

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.[10] Second, we use the single-period lagged variable $\ln Function\_turbulence_{sp-1}$. By using predetermined variables in the regression, we can establish whether or not function turbulence is predictive of style turbulence. Third, we include a quadratic term, $(\ln Function\_turbulence_{sp-1})^2$, because we are positing a curvilinear relationship between the effect of function turbulence and the effect of style turbulence. The full specification is

$$\ln T_{sp} = c_i + \beta_1(\ln Function\_turbulence_{sp-1}) + \beta_2[(\ln Function\_turbulence_{sp-1})^2] + \beta_3(Period) + \beta_4(\ln Style\_activity_{sp}) + \beta_5(\ln Category\_activity_{ip}) + \varepsilon. \quad (1)$$

[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.

We are mainly interested in the coefficients β_1, β_2, and β_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 (9,352 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 $\ln Function\_turbulence_{sp-1}$ 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. Figure 9 plots the marginal effects of $\ln Function\_turbulence_{sp-1}$ (estimated semiparametrically);[11] the graph confirms our hypothesized U-shaped relation.

[11] To perform this estimation, we first segment $\ln Function\_turbulence_{sp-1}$ into seven discrete "buckets". More specifically, we create indicator variables K_j, $j = 0, \ldots, 5$, that take the value 1 when $\ln Function\_turbulence_{sp-1}$ is in the range $[j, j+1)$ and otherwise take the value 0; we put $K_6 = 1$ when $\ln Function\_turbulence_{sp-1}$ is greater than 6 (K_6 corresponds to about 1% of the data). Then we estimate the effects of each bucket on our dependent variable while controlling for the other variables in Equation 1. This approach allows us to model any arbitrary (stepwise) relation between the variables $\ln Function\_turbulence_{sp-1}$ and $\ln Style\_turbulence_{sp}$.

Table 2: Regression model predicting style turbulence (*N* = 28,483 style-period observations)

Variable	Function Only	Time Only	Full Model	Semiparametric
ln $Function\_turbulence_{sp-1}$	–4.62*** (0.37)		–4.70*** (0.37)	—
(ln $Function\_turbulence_{sp-1}$)^2	0.81*** (0.09)		0.81*** (0.09)	—
Period		0.70** (0.24)	1.30*** (0.24)	1.46*** (0.24)
ln $Style\_activity_{sp}$	35.41*** (0.28)	34.92*** (0.32)	35.48*** (0.28)	35.58*** (0.29)
ln $Category\_activity_{ip}$	5.65*** (0.40)	2.21* (0.89)	1.50 (0.89)	1.33 (0.89)
Product category fixed effects	Yes	Yes	Yes	Yes
ln $Function\_turbulence_{sp-1}$ dummies	No	No	No	Yes
R^2	0.75	0.75	0.75	0.76
F-statistic	631***	428***	617***	541***

Notes: Standard errors (in parentheses) are clustered by style. Coefficients and standard errors are scaled 100×.
$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

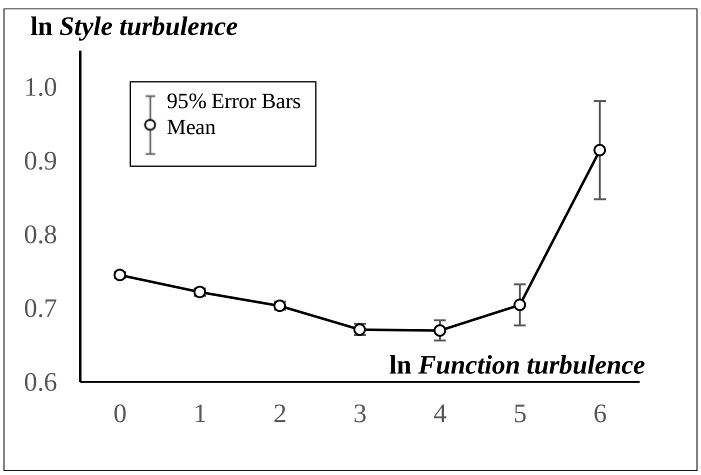

Figure 9: Marginal effects of function turbulence on style turbulence

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; 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 $\ln Function\ activity = \ln \sum_{u \in \boldsymbol{U}_{sp}} M_{up}$, 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-shape 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 $\{x_{sp}^1, \dots, x_{sp}^5\}$), 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, $x_{sp}^{i+1} = x_{sp}^i + u + \varepsilon$. The unpredictability of this process is captured by the error term (ε). We can estimate the standard error of ε by taking first differences and then calculating the sample standard deviation: $S_{sp}(\varepsilon) = \sqrt{\mathrm{Var}(\Delta x_{sp})}$ (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., $\ln Style\ turbulence$) 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.

Table 3: Results of selected robustness tests (*N* = 28,483 style-period observations)

Variable	Model 1	Model 2	Model 3	Model 4	Model 5
ln $Function\_turbulence_{sp-1}$	-10.13^{***} (1.00)	-4.51^{***} (0.38)	-2.81^{***} (0.31)	-5.32^{***} (0.45)	-4.93^{***} (0.38)
(ln $Function\_turbulence_{sp-1})^2$	1.14^{***} (0.12)	0.77^{***} (0.09)	0.60^{***} (0.07)	0.83^{***} (0.09)	0.87^{***} (0.09)
Period	1.27^{***} (0.24)	—	1.11^{***} (0.30)	1.34^{***} (0.24)	1.27^{***} (0.24)
ln $Style\_activity_{sp}$	35.52^{***} (0.28)	35.47^{***} (0.29)	44.31^{***} (0.29)	34.98^{***} (0.25)	35.40^{***} (0.29)
ln $Function\_activity_{sp-1}$	2.54^{***} (0.35)				
ln $Category\_activity_{ip}$	1.26 (0.88)	—	2.43^{*} (1.09)	1.50 (0.88)	1.50 (0.88)
ln $Style\_turbulence_{sp-1}$				3.39^{***} (0.77)	
Category fixed effects	Yes	No	Yes	Yes	Yes
Category-period fixed effects	No	Yes	No	No	No
R^2	0.76	0.76	0.77	0.75	0.75
F-statistic	590^{***}	182^{***}	970^{***}	677^{***}	588^{***}

Notes: Standard errors (in parentheses) are clustered by style. Coefficients and standard errors are scaled 100×.
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

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 themselves contain the names of inventors, which are needed to create our instrumental variables. We rely also on the disambiguation of names into unique inventor IDs by Lai et al. (2014). We start by creating two variables for each period p of a utility category u. The first variable is $Arrival$, defined as the number of unique inventors observed in category u for period p but not for $p-1$. The second is $Departure$, 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 $\ln Function\_turbulence$.[12] 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 $\ln Arrival$ and 8,053 for $\ln Departure$—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 four-year and ten-year instead of five-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

[12] 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 (cf. 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.

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 tool kits 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- 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?—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.

REFERENCES

Alexander, C. 1964. *Notes on the Synthesis of Form*. Harvard University Press.

Anderson, P., M. L. Tushman. 1990. Technological discontinuities and dominant designs: A cyclical model of technological change. *Adm. Sci. Q.* **35** 604–633.

Aral, S., D. Walker. 2011. Viral Product Design : A Randomized Trial of Peer Influence in Networks. *Manage. Sci.* **57**(9) 1623–1639.

Armstrong, J. S. 2001. *Principles of Forecasting: A Handbook for Researchers and Practitioners* J. S. Armstrong, ed. Springer Science & Business Media.

Baldwin, C. Y., K. B. Clark. 2000. *Design rules: The power of modularity*. MIT Press.

Barlas, Y. 1996. Formal aspects of model validity and validation in system dynamics. *Syst. Dyn. Rev.* **12**(3) 183–210.

Bloch, P. H. 1995. Seeking the Ideal Form: Product Design and Consumer Response. *J. Mark.* **59**(3) 16.

Bloch, P. H. 2011. Product Design and Marketing: Reflections After Fifteen Years. *J. Prod. Innov. Manag.* **28**(3) 378–380.

Bourgeois, L., K. Eisenhardt. 1988. Strategic decision processes in high velocity environments: four cases in the microcomputer industry. *Manage. Sci.* **34**(7) 816–835.

Brown, T. 2009. *Change by Design*. Harper Collins.

Cachon, G. P., T. Randall, G. M. Schmidt. 2007. In Search of the Bullwhip Effect. *Manuf. Serv. Oper. Manag.* **9**(4) 457–479.

Cameron, A. C., P. K. Trivedi. 2009. *Microeconometrics using Stata*. College Station, TX, Stata Press.

Cappetta, R., P. Cillo, A. Ponti. 2006. Convergent designs in fine fashion: An evolutionary model for stylistic innovation. *Res. Policy* **35**(9) 1273–1290.

Chan, C.-S. 2000. Can style be measured? *Des. Stud.* **21**(3) 277–291.

Chung, F. 1997. *Spectral Graph Theory*. American Mathematical Society.

Church, R. M., P. Guilhardi. 2005. A Turing test of a timing theory. *Behav. Processes* **69**(1) 45–58.

Cillo, P., G. Verona. 2008. Search Styles in Style Searching: Exploring Innovation Strategies in Fashion Firms. *Long Range Plann.* **41**(6) 650–671.

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ, Lawrence Erlbaum Associates.

Dell'Era, C., A. Marchesi, R. Verganti. 2010. Mastering Technologies in Design-Driven Innovation. *Res. Technol. Manag.* **53**(2) 12–23.

Dess, G. G., D. W. Beard. 1984. Dimensions of Organizational Task Environments. *Adm. Sci. Q.* **29**(1) 52–73.

Dougherty, D. 1990. Understanding new markets for new products. *Strateg. Manag. J.* **11**(Special Issue: Corporate Entrepreneurship) 59–78.

Dougherty, D. 2001. Reimagining the Differentiation and Integration of Work for Sustained Product Innovation. *Organ. Sci.* **12**(5) 612–631.

Du Mont, J. J., M. D. Janis. 2011. The Origins of American Design Protection. *Indiana Law J.* **88** 11–18.

The Economist. 2012. Apple v Samsung: iPhone, uCopy, iSue. Sept. 1st 2012.

Eisenhardt, K. 1989. Making fast strategic decisions in high-velocity environments. *Acad. Manag. J.* **32**(3) 543–576.

Eisenhardt, K., B. Tabrizi. 1995. Accelerating adaptive processes: Product innovation in the global computer industry. *Adm. Sci. Q.* **40**(1) 84–110.

Eisenman, M. 2013. Understanding Aesthetic Innovation in the Context of Technological Evolution. *Acad. Manag. Rev.* **38**(3) 332–351.

Erat, S., V. Krishnan. 2012. Managing delegated search over design spaces. *Manage. Sci.* **58**(3) 606–623.

Ferguson, C. J. 2009. An effect size primer: A guide for clinicians and researchers. *Prof. Psychol. Res. Pract.* **40**(5) 532–538.

Fine, C. H. 1998. *Clockspeed: Winning industry control in the age of temporary advantage*. Basic Books.

Fleming, L., O. Sorenson. 2004. Science as a map in technological search. *Strateg. Manag. J.* **25**(89) 909–928.

Geman, D., S. Geman, N. Hallonquist, L. Younes. 2015. Visual Turing test for computer vision systems. *Proc. Natl. Acad. Sci. U. S. A.* **112**(12) 3618–23.

Godart, F. 2012. *Unveiling fashion: business, culture, and identity in the most glamorous industry*. Palgrave Macmillan.

Goldstone, R. L. 1994. The role of similarity in categorization: providing a groundwork. *Cognition* **52**(2)

125–57.

Granger, C. W. J. 1969. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **37**(3) 424–438.

Hall, B. H., A. B. Jaffe, M. Trajtenberg. 2005. Market value and patent Citations. *RAND J. Econ.* **86** 2418–2427.

Hargadon, A. B., Y. Douglas. 2001. When innovations meet institutions: Edison and the design of the electric light. *Adm. Sci. Q.* **46**(3) 476–501.

Henderson, R., K. B. Clark. 1990. Architectural innovation: the reconfiguration of existing product technologies and the failure of established firms. *Adm. Sci. Q.* 9–30.

Hildebrand, C., G. Häubl, A. Herrmann, J. R. Landwehr. 2013. When Social Media Can Be Bad for You: Community Feedback Stifles Consumer Creativity and Reduces Satisfaction with Self-Designed Products. *Inf. Syst. Res.* **24**(1) 14–29.

Hoegg, J., J. W. Alba. 2011. Seeing Is Believing (Too Much): The Influence of Product Form on Perceptions of Functional Performance. *J. Prod. Innov. Manag.* **28**(3) 346–359.

Holland, J. H., K. J. Holyoak, R. E. Nisbett, P. R. Thagard. 1986. *Induction: Processes of inference, learning, and discovery*. Cambridge, MA, MIT Press.

Hölttä-Otto, K., V. Tang, K. Otto. 2008. Analyzing module commonality for platform design using dendrograms. *Res. Eng. Des.* **19**(2-3) 127–141.

Jones, B. F. 2009. The burden of knowledge and the “death of the renaissance man”: Is innovation getting harder? *Rev. Econ. Stud.* **76**(1) 283–317.

Jupp, J., J. Gero. 2006. Visual style: Qualitative and context-dependent categorization. *AIE EDAM Artif. Intell. Eng. Des. Anal. Manuf.* **20**(3) 247–266.

Kannan, R., S. Vempala, A. Vetta. 2004. On clusterings: Good, bad and spectral. *J. ACM* **51**(3) 497–515.

Kessler, M. M. 1963. Bibliographic coupling between scientific papers. *Am. Doc.* **14**(1) 10–25.

Kornish, L. J., K. T. Ulrich. 2011. Opportunity spaces in innovation: Empirical analysis of large samples of ideas. *Manage. Sci.* **57**(1) 107–128.

Kreuzbauer, R., A. J. Malter. 2005. Embodied Cognition and New Product Design: Changing Product Form to Influence Brand Categorization*. *J. Prod. Innov. Manag.* **22**(2) 165–176.

Krippendorff, K., R. Butter. 1993. Where Meanings Escape Functions. *Des. Manag. J.* **4**(2) 30–37.

Landwehr, J. R., A. A. Labroo, A. Herrmann. 2011. Gut Liking for the Ordinary: Incorporating Design Fluency Improves Automobile Sales Forecasts. *Mark. Sci.* **30**(3) 416–429.

Landwehr, J. R., D. Wentzel, A. Herrmann. 2013. Product Design for the Long Run: Consumer Responses to Typical and Atypicai Designs at Different Stages of Exposure. *J. Mark.* **77**(5) 92–107.

Lei, Z., B. Wright. 2009. Why weak patents? Rational ignorance or pro-'customer' tilt. *Univ. California, Berkeley*.

Li, G.-C., R. Lai, A. D. Amour, et al. 2014. Disambiguation and Co-authorship Networks of the U . S . Patent Inventor Database (1975-2010). *Res. Policy* **43**(6) 941–955.

Loch, C. H., C. Terwiesch. 1998. Communication and Uncertainty in Concurrent Engineering. *Manage. Sci.* **44**(8) 1032–1048.

Maeda, J. 2015. Have Designers Killed the Engineering Star? Interview by Emily Chang. *Bloom. Bus.* Available at: http://www.bloomberg.com/news/videos/2015-05-19/the-growing-importance-of-designers-in-silicon-valley [Accessed June 7, 2015].

McCormack, J., J. Cagan, C. Vogel. 2004. Speaking the Buick language: capturing, understanding, and exploring brand identity with shape grammars. *Des. Stud.* **25** 1–29.

McNamara, G., P. M. Vaaler, C. Devers. 2003. Same as it ever was: the search for evidence of increasing hypercompetition. *Strateg. Manag. J.* **24**(3) 261–278.

Mendelson, H., R. R. Pillai. 1999. Industry Clockspeed: Measurement and Operational Implications. *Manuf. Serv. Oper. Manag.* **1**(1) 1–20.

Mervis, C. B., E. Rosch. 1981. Categorization of natural objects. *Annu. Rev. Psychol.* **32** 89–115.

Miller, C., W. Glick. 2006. Assessing the external environment: an enrichment of the archival tradition.

Res. Methodol. Strateg. Manag. **3** 97–122.

Molotch, H. 2004. *Where stuff comes from: How toasters, toilets, cars, computers and many other things come to be as they are*. Routledge.

Munro, T. 1946. Style in the arts: a method of stylistic analysis. *J. Aesthet. Art Crit.* **5**(2) 128–158.

Nakagawa, S., I. C. Cuthill. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Rev. Camb. Philos. Soc.* **82**(4) 591–605.

Ng, A., M. Jordan, Y. Weiss. 2002. On spectral clustering: Analysis and an algorithm. *Adv. Neural Inf. Process. Syst.* **2** 849–856.

Noble, C. H. 2011. On Elevating Strategic Design Research. *J. Prod. Innov. Manag.* **28**(3) 389–393.

Oppy, G., D. Dowe. 2016. The Turing Test. E. N. Zalta, ed. *Stanford Encycl. Philos.*

Orth, U. R., K. Malkewitz. 2008. Holistic Package Design and Consumer Brand Impressions. *J. Mark.* **72**(3) 64–81.

Paolacci, G., J. Chandler, P. Ipeirotis. 2010. Running Experiments on Amazon Mechanical Turk. *Judgm. Decis. Mak.* **5**(5) 411–419.

Porac, J., H. Thomas. 1990. Taxonomic mental models in competitor definition. *Acad. Manag. Rev.* **15**(2) 224–240.

Pugliese, M., J. Cagan. 2002. Capturing a rebel: modeling the Harley-Davidson brand through a motorcycle shape grammar. *Res. Eng. Des.* **13** 139–156.

Ravasi, D., G. Lojacono. 2005. Managing design and designers for strategic renewal. *Long Range Plann.* **38**(1) 51–77.

Ravasz, E., A.-L. Barabási. 2003. Hierarchical organization in complex networks. *Phys. Rev. E* **67**(2) 026112.

Rindova, V. P., A. P. Petkova. 2007. When Is a New Thing a Good Thing? Technological Change, Product Form Design, and Perceptions of Value for Product Innovations. *Organ. Sci.* **18**(2) 217–232.

Rosch, E. 1978. Principles of categorization. *Cogn. Categ.* Hilsdale, NJ, Erlbaum, 27–48.

Rosch, E., C. Mervis. 1975. Family resemblances: Studies in the internal structure of categories. *Cogn. Psychol*. **605** 573–605.

Rubera, G., C. Droge. 2013. Technology versus design innovation's effects on sales and tobin's Q: The moderating role of branding strategy. *J. Prod. Innov. Manag.* **30**(3) 448–464.

Sargent, R. G. 1999. Validation and verification of simulation models. *Proc. 1999 Winter Simul. Conf.* 39–48.

Sawyer, A., A. Ball. 1981. Statistical power and effect size in marketing research. *J. Mark. Res.* **18**(3) 275–290.

Shieber, S. M. 2007. The Turing test as interactive proof. *Nous* **41**(4) 686–713.

Simon, H. A. 1969. The sciences of the artificial. *Cambridge, MA.*

Solon, G. 1992. Intergenerational income mobility in the United States. *Am. Econ. Rev*. **82**(3) 393–408.

Stacey, M. 2006. Psychological challenges for the analysis of style. *Artif. Intell. Eng. Des. Anal. Manuf.* **20**(Special Issue on Style) 1–40.

Stiny, G. 1980. Introduction to shape and shape grammars. *Environ. Plan. B* **7**(3) 343–351.

Stock, J. H., J. H. Wright, M. Yogo. 2002. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *J. Bus. Econ. Stat.* **20**(4) 518–529.

Sullivan, L. H. 1896. The tall office building artistically considered. *Lippincott's Mag.* **57**(3) 406.

Teece, D., G. Pisano, A. Shuen. 1997. Dynamic capabilities and strategic management. *Strateg. Manag. J.* **18**(7) 509–533.

Terwiesch, C., C. H. Loch. 2004. Collaborative Prototyping and the Pricing of Custom-Designed Products. *Manage. Sci.* **50**(2) 145–158.

Terwiesch, C., K. T. Ulrich. 2009. *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Harvard Business Press.

Thomke, S. H. 1998. Managing Experimentation in the Design of New Products. *Manage. Sci.* **44**(6) 743–762.

Turing, A. M. 1950. Computing machinery and intelligence. *Mind* 433–460.

Tushman, M., P. Anderson. 1986. Technological discontinuities and organizational environments. *Adm. Sci. Q.* **31**(3) 439–465.

Ulrich, K. T. 2011. *Design: Creation of artifacts in society*. University of Pennsylvania.

USPTO. 2005. *Handbook of classification*. USPTO.

USPTO. 2006. *Manual of Patent Examining Procedures*. USPTO.

Utterback, J. M. 1996. *Mastering the dynamics of innovation*. Harvard Business Press.

Utterback, J. M., W. J. Abernathy. 1975. A dynamic model of process and product innovation. *Omega* **3**(6) 639–656.

Van Horn, R. L. 1971. Validation of simulation results. *Manage. Sci.* **17**(5) 247–258.

Verganti, R. 2006. Innovating Through Design. *Harv. Bus. Rev.* **84**(12) 114.

Veryzer, R. W. 2005. The Roles of Marketing and Industrial Design in Discontinuous New Product Development. *J. Prod. Innov. Manag.* **22**(1) 22–41.

Wooldridge, J. 2010. *Econometric analysis of cross section and panel data*. Cambridge, MA, MIT Press.

Ziliak, S. T., D. N. McCloskey. 2004. Size matters: The standard error of regressions in the American Economic Review. *J. Socio. Econ.* **33**(5) 527–546.