2021

Order matters: How altering the sequence of performance events shapes perceived quality formation

Owen Parker

Ke Gong

Rachel Mui

Varkey K. Titus Jr.

Jiancheng Du

See next page for additional authors

Authors

Owen Parker, Ke Gong, Rachel Mui, Varkey K. Titus Jr., Jiancheng Du, and Gyebi Kwarteng

Order matters: How altering the sequence of performance events shapes perceived quality formation

Owen Parker,[1] Ke Gong,[2] Rachel Mui,[3] Varkey Titus,[4] Jiancheng Du,[5] and Gyebi Kwarteng[6]

1 University of Texas at Arlington, Department of Management, College of Business
2 Iowa State University, Ivy College of Business
3 Kansas State University, Department of Management, College of Business Administration
4 University of Nebraska-Lincoln, Department of Management, College of Business
5 Old Dominion University, Strome College of Business
6 Oklahoma State University, Department of Management, Spears School of Business

Corresponding author — O. Parker, 701 S. West Street, Arlington, TX 76019
E-mail addresses: owen.parker@uta.edu (O. Parker), kegong@iastate.edu (K. Gong), rachelmui@ksu.edu (R. Mui), vtitusjr2@unl.edu (V. Titus), jduo16@odu.edu (J. Du), gyebi.kwarteng@okstate.edu (G. Kwarteng).

Abstract

Reputation research often employs rankings which combine both the prominence and perceived quality dimensions of reputation. Though this approach has merit, it neglects nuances in the formation of perceived firm quality – i.e., how stakeholders

Published in *Journal of Business Research* 126 (2021), pp. 48–63.
doi:10.1016/j.jbusres.2020.12.043

Submitted 28 April 2020; revised 15 December 2020; accepted 16 December 2020; published 2 January 2021.

perceive a firm's capabilities. Since perceptions are influenced by how information is presented, we posit that the patterns of a firm's performances – their order and interval – explain variance in perceived quality beyond valence (absolute performance level), alone. We employ two experiments and an archival study to manipulate product ratings and collect perceived quality scores (experimentally), and use trajectory of performance outcomes to predict market valuation as a perceived quality proxy (archivally). Results suggest that while valence matters most for a firm's perceived quality, presenting identical performance events with distinct orders and intervals changes perceived quality impressions, at least until new information is presented. We enumerate our findings and outline areas for future research on stakeholder perceptions.

Keywords: Perceptions, Reputation, Information processing, Sequences, Temporality, Stakeholders

1. Introduction

Stakeholder perceptions about the merits of an organization's achievements or operations are important for numerous reasons. Broadly speaking, a range of fields across the organization sciences deal with these perceptions, whether they call them market expectations (e. g., Fried & Givoly, 1982), subjective quality (e.g., Mitra & Golder, 2006), or simply 'stakeholder perceptions' (Zavyalova, Pfarrer, Reger, & Shapiro, 2012). In the social evaluations realm, research indicates that stakeholders' impression of the firm's *perceived quality*—that is, the degree to which stakeholders positively evaluate the firm for the quality of its offerings, operations, or achievements—can increase the price premium a firm can extract for its products or services (e.g., Rindova, Williamson, Petkova, & Sever, 2005) and reduce uncertainty about the firm's prospects and capabilities (e.g., Shapiro, 1982). A firm's perceived quality can even shape the firm's strategic decisions, such as the rate of product introduction (e.g., Parker, Krause, & Covin, 2017). In the reputation literature, perceived quality was first conceptualized—alongside a firm's "prominence"—as one critical component of a firm's reputation (e.g., Rindova et al., 2005). The literature on perceived quality is adjacent to, but distinct from, the literature assessing overall reputation as a combination of both prominence and perceived quality, such as via "Fortune's *Most Admired Companies*" rankings (e.g., Haleblian, Pfarrer, & Kiley, 2017; Pfarrer, Pollock, & Rindova, 2010). Perceived

quality research is intertwined with the later notion that firms are most precisely conceptualized as "being known for something" (e.g., Lange, Lee, & Dai, 2011), and that this "something" can be broadly characterized as the firm's quality or capability on some specific dimension of performance (e.g., Jensen, Kim, & Kim, 2012). Yet while perceived quality is undoubtedly an important concept, our understanding of how it forms, in the first place, is incomplete.

At the product level, Mitra and Golder (2006) examine how a particular product's "objective" quality influences the perceived quality of that product, and how it takes time for that evidence to be reflected in stakeholders' objective perceptions. However, macro-level research on the topic lags behind in terms of such granularity. The scant research that does exist focuses almost entirely on how perceived quality is shaped by a firm's performance valence—i.e., how "good" or "bad" quality-related performances are in the eyes of stakeholders (e.g., Rhee & Haunschild, 2006; Parker et al., 2017). Interestingly, it has long been argued that perceived quality is shaped "by the signals that organizations send when they make strategic choices about the resources deployed in producing products and services" (Rindova et al., 2005: 1034). However, the other facets of these signals, such as their interval and order, have been eschewed in favor of valence, alone.

Consider, for example, that different firms can have the same average valence of performance over a certain period, yet the order of specific performance incidents can vary: a particular firm might start with strong performance and then rapidly worsen in future performances, while another might begin with a poor performance but slowly and steadily improve over time. Although the firm-level research on perceived quality says little about the implications of such a scenario, the psychology literature on judgment and information processing indicates that stakeholders will perceive and process such disparate sequences differently (Hohle & Teigen, 2015), leading to distinct perceived quality evaluations in the minds of those stakeholders. The tendency of past perceived quality research to focus on performance valence ignores the potential influence of these performance sequences, despite the fact that it is plausible that stakeholders' perceived quality of the firm are influenced by more than just the rational, cumulative evidence about the valence of a firm's quality

achievements. We examine the perceived quality formation process beyond that which is driven by performance valence signals indicating "good" or "bad" performance over time. Though this concept pervades the organization sciences, for the sake of clarity and specificity we root our conceptualization of perceived quality primarily in the way it is characterized in the social evaluations stream (e.g., Bitektine, 2011; Ravasi, Rindova, Etter, & Cornelissen, 2018; Rhee & Haunschild, 2006; Rindova et al., 2005).

In reality, performance evidence often comes in bits and pieces, and we contend that the piecemeal nature of this evidence shapes perceptions as stakeholders observe this evidence and incrementally form an impression as to whether the firm is headed in a certain direction or whether those signals are steady and consistent enough to be "meaningful indicators of quality". Since perceived quality is fundamentally a subjective impression, it is important to consider the extent to which perceived sequences or patterns in performance events might shape a firm's perceived quality in the minds of stakeholders. Scholars have also argued this for the overall concept of reputation, regardless of how it is conceptualized (e.g., Ravasi et al., 2018; Schmidt, 2018).

As described in the example scenarios mentioned above, as well as research on information processing, sequences of performance evidence can take on different *orders* (e.g., worsening, improving, or flat) and different *intervals* (e.g., rapid or gradual), and these aspects characterize an array of performance types on which stakeholders' perceived quality of the firm are based. Perceived quality is similarly shaped by the order and interval of that performance evidence as stakeholders perceive a certain "trajectory" about where the firm is headed.

This prompts several key questions. First, how do the order and interval of performance events shape stakeholders' perceived quality of the firm beyond the valence of those events, alone. Second, do order and interval sometimes outweigh valence in influencing these impressions? These questions form the foundation for our subsequent argumentation and testing. We forgo formal hypotheses because existing literature does not provide a clear basis upon which we can defensibly hypothesize how order and interval influence stakeholders' perceived quality of the firm.

We therefore examine these questions using two experimental studies and one archival study that each incrementally add insight to our research question. Our findings provide several important insights about the perceived quality formation process, and we draw on existing theory and argumentation to explain these findings. Our study represents a first effort in research to address how the temporality of performance signals influence stakeholders' perceived quality of the firm.

Our first experiment indicates that trajectory can sometimes—albeit very rarely—*override* valence, such as when a poorer average performer exhibits a pattern of dramatically improving performance. This is echoed by our results from our second study based on archival data, which indicate that improving orders have a better perceived quality outcome than worsening orders, even when we control for numerous extraneous factors that might otherwise shape this relationship. The implication here is that managers can attend to these trajectory effects when scheduling product releases or deciding which venture to pursue first. This suggests that the perceived quality of a firm can be strategically shaped by a series of actions with a carefully designed pattern of expected quality signals. Managers might order products to be introduced, for instance, in a sequence based on the amount of resources invested.

Second, we find that interval can exacerbate the order effect, such that performance patterns that unfold steadily over time have different implications than patterns that emerge rapidly. Additionally, our experiments indicate that these differences in stakeholders' perceived quality of the firm *erode* when new information is presented, suggesting that the effects of recency override the effects of different orders or intervals in performance sequences. In other words, when respondents are shown equal performance valence in the final period after seeing a random sequence of improving/worsening and rapid/gradual performance events, the differences in their perceived quality impressions across these scenarios disappear. Practically speaking, this suggests that managers may find that although performance sequences do seem to matter for perceived quality impressions, unless these sequences are perpetuated, the effects are generally short-lived.

Our research contributes to the literature in several ways. First, we contribute to the conversation on how meaning-making among stakeholders influences perceived quality formation. Some research has addressed how perceived quality is rooted in prior performances (e.g., Rindova et al., 2005; Rhee & Haunschild, 2006; Parker et al., 2017), but we focus not only on the valence of those performances but also their order and interval.

Relatedly, we contribute to research on social evaluations. While prior research shows that managers seem to bundle together activities for which they anticipate a backlash (e.g., Titus, Parker, & Bass, 2018), research has yet to identify the consequences of this strategy—i.e., whether it is advantageous vis-`a-vis protecting the firm's perceived quality or other intangible assets. A firm's perceived quality forms in the minds of stakeholders who are prone to the same biases discussed in the judgment and decision making literature (e.g., Chaiken, 1980; Wyer Jr, 2003). Notably, there is almost no discussion of order and interval effects—or, indeed, any temporal effects—for any of the social evaluation constructs in the prior literature (for an exception, see Carter & Ruefli, 2006). By examining the microfoundations of perceived quality, we also shed light on a critical contributor to reputation (e.g., Rindova et al., 2005; Parker, Krause, & Devers, 2019). For example, the results of our third study indicate that a "rapid decrease" of performance valence damages perceived quality less than a gradual, protracted decrease of performance valence. This underscores the merit of "ripping off the Band-Aid" (e.g., Titus et al., 2018: 637) in protecting a firm from the hazards of negative stakeholder perceptions, and is an initial step in examining the temporal effects of performance events on stakeholder perceptions.

Finally, our work demonstrates the usefulness of experiments in the social evaluations literature, an approach which has been neglected relative to archival methods (for a recent exception to this, see Schmidt, 2018). We believe that our multi-method approach in examining this important question around perceived quality impressions illustrates the ways in which scholars not accustomed to using experimental approaches can integrate aspects of this method into their empirical analyses.

2. Theoretical background

2.1. Perceived quality's role in the broader literature

We define perceived quality as the degree to which stakeholders positively evaluate an organization for the quality of its offerings, operations, or achievements. Reputation has been conceptualized as consisting of two core elements: a firm's prominence and its perceived quality (e.g., Rindova et al., 2005). Importantly, while some scholars have studied reputation for quality (e.g., Dierickx & Cool, 1989; Parker et al., 2017; Rhee & Haunschild, 2006), perceived quality is not exclusively a reputation for quality, but also includes the quality of the firm's efforts, such as in meeting stakeholder expectations with respect to other outcomes or processes (e.g., profitability, market share). This is consistent with how scholars characterize perceived quality as a dimension of general reputation vis-`a-vis "being good" at what you do (e.g., Rindova et al., 2005), and as a cluster of potential paths toward "being known *for something*" (e.g., Lange et al., 2011, emphasis added).

Although many studies have examined reputation from the perspective of both widespread notoriety or prominence and achievement on e.g., financial performance (Fortune's "Most Admire Companies" ranking is the most frequently used reputation proxy for reputation, see Pollock, Lashley, Rindova, and Han (2019)), perceived quality *by itself* is also worth examining in closer detail. That is, studies that examine perceived quality demonstrate its importance for numerous considerations including pricing (e.g., Rindova et al., 2005), product introduction decisions (e.g., Parker et al., 2017), and product recalls (e.g., Rhee & Haunschild, 2006). However, we acknowledge the myriad forms that perceived quality can take in our definition of perceived quality as the extent of stakeholders' positive evaluations of the organization's "offerings, *operations, or achievements*". A firm's "perceived quality" is an assessment of the *firm* and not just its "product quality", which involves subjective assessments about specific product or service offerings (e.g., Mitra & Golder, 2006). More broadly speaking, perceived quality can also represent how well stakeholders believe the firm is doing on a variety of other dimensions that the audience may prioritize, such as innovation output or employee satisfaction (e.g., Jensen et al., 2012).

Leaving aside this likely imprecision of the word "quality", considered by itself, a firm's perceived quality is shaped "by the signals that organizations send when they make strategic choices about the resources deployed in producing products and services" (Rindova et al., 2005: 1034). That is, is the organization good at what it purports to do? In that vein, stakeholders' inferences are critically important to the formation of perceived quality, and we turn our attention to those processes.

2.2. Stakeholder inference of meaning from performance patterns

Perceived quality is often rooted in stakeholders' subjective interpretations of the firm's performance (e.g., Rindova et al., 2005). This suggests that perceived quality is a product of stakeholders' information processing, and as such, it is critically important to develop a deeper scholarly understanding of how these processes influence the formation of these impressions.

According to the information processing literature, individuals are known to attempt to derive meaning from patterns of information (Hogarth & Einhorn, 1992; Tversky & Kahneman, 1974). For instance, individuals may perceive a series of successive positive events as indicative of a "streak" which they may believe increases the likelihood of yet another favorable outcome (e.g., Gilovich, Vallone, & Tversky, 1985). Even brief sequences of information can generate these kinds of perceptual biases as individuals seek to make sense of the available data. As more information becomes available, any preconceived expectations may be reinforced (e.g., Hogarth & Einhorn, 1992) as individuals are more likely to selectively attend to information that confirms those expectations (Griffin & Tversky, 1992; Rabin & Schrag, 1999).

2.2.1. Higher valence is better

In general, the micro-cognitive perspective on social judgments such as perceived quality contends that these impressions form through information processing as individuals become exposed to incrementally more evidence about the firm. In this view, such judgments amount to "an ongoing reevaluation" (Ravasi et al., 2018), whereby stakeholder-observers update their impressions of the firm's perceived quality as

more evidence becomes available. Some researchers have examined social judgments in terms of duration of performance (e.g., Brandts & Figueras, 2003) and the belief formation process (Fischer & Reuber, 2007).

However, most of this prior research implicitly takes a binary, valence-focused approach, assuming that performances which are consistently good engender favorable impressions of perceived quality, while performances that are consistently bad engender unfavorable impressions of perceived quality. While this view of perceived quality formation has served as a good starting point for theorizing, performance patterns are rarely solely favorable or solely unfavorable. Over time, performances may take on patterns with different order and interval characteristics that may have distinct and important influences on perceived quality formation, because of how one processes these characteristics holistically.

2.2.2. Order and interval differences

Over a certain period, the same average level of performance valence can be comprised of events which have distinct order and interval traits, and these may shape stakeholder perceptions beyond average performance valence, alone. Information processing research has theorized how individuals might judge a firm based on the firm's sequence of performance (Bitektine, 2011; Ravasi et al., 2018), and scholars have emphasized the need for research to further determine how sequences can impart meaning based on the valence, trajectory, and interval of signals observed (Connelly, Certo, Ireland, & Reutzel, 2011). Some research indicates that a decline in performance can erode social judgments adjacent to perceived quality (e.g., reputation) because of the valence of those performances (e.g., Basdeo, Smith, Grimm, Rindova, & Derfus, 2006; Love & Kraatz, 2009), and that smaller declines may not erode those impressions as much (Rhee & Valdez, 2009).

Valence, or the absolute level of average performances that is usually classified as either "good" or "bad", is often the most intuitive performance characteristic in that higher valence performances are generally associated with more favorable perceived quality impressions (Rindova et al., 2005). With respect to ***order***, a sequence of performances can be ordered to represent an improving or worsening

trajectory, or may lack any discernible trajectory—e.g., a flat line or fluctuating performance. ***Interval*** describes the rate at which the performance signals are revealed, e.g., in rapid succession or gradually over a long period. Any of these characteristics of performance sequences could influence the process of perceived quality formation in the minds of stakeholders, and yet we know very little about how this might unfold or which element might take precedence.

2.3. Does order matter?

Social judgment researchers argue and show that posting favorable performance prompts high stakeholder expectations, which in turn means that poor future performances will be more threatening to the firm (Parker et al., 2019; Rhee & Haunschild, 2006; Zavyalova et al., 2012; Zavyalova, Pfarrer, Reger, & Hubbard, 2016). Rhee and Haunschild (2006) made this argument with respect to product recalls by Toyota, a firm which—at the time of their recall of cars with stuck accelerators—had industry-leading levels of perceived quality in the eyes of stakeholders. Toyota and other firms in such a position, they argue, are disproportionately damaged by such crises when this constitutes what we might call a "fall from grace". Parker et al. (2019) offer a theoretical exposition on how possessing a good reputation—for attributes such as product quality or other capabilities—binds the hands of managers as they contemplate their strategic discretion, because these decision makers fear missteps will disappoint stakeholders.

Though perceived quality does not comprise the entirety of a firm's reputation, as noted previously, it has been characterized as one component of reputation (e.g., Rindova et al., 2005; Lange et al., 2011); more importantly, all of these studies seem to indicate that perhaps an initial success followed by less favorable evidence will generate worse perceived quality in the eyes of stakeholders than if the order were different. That is, an order which seems to suggest a "worsening trajectory" may motivate unfavorable perceived quality impressions, while one which seems to indicate an "improving trajectory" may engender more favorable perceived quality impressions.

2.4. Does interval matter?

Although the prior literature has predominantly focused on valence differences vis-`a-vis perceived quality formation, there are hints of what we might expect from order and interval differences in terms of their role in shaping perceived quality impressions.

2.4.1. The "slower means salient" view

Some recent evidence suggests that managers may hasten firm activities that are expected to generate negative scrutiny—especially when attention on the firm is already unfavorable—ostensibly because they fear the hazard of delaying more than the hazard of "ripping off the Band-Aid," i.e., attracting substantial negative attention in the short-run (e.g., Titus et al., 2018). In this view, stakeholders are conceptualized as having limited attention and a ceiling on their indignation, such that once the firm is already facing the burden of negative attention, it is perhaps advantageous to go ahead and engage in what these authors term "scrutiny bundling" behaviors, lumping all of these hazardous actions (e.g., oil drilling, deforestation, revelation of damning information) into a brief period and hoping for a steady recovery in the future.

2.4.2. The "speed is symbolic" view

At the level of individual performance, Gilovich et al. (1985) present compelling evidence of the "hot hand" or "streak shooting" fallacies in professional basketball. The authors show that despite the common perception that a sequence of numerous successive "shots scored" increases the likelihood that the next shot will also score, there is no evidence of such predictions holding true. As the authors note, the "belief in the hot hand and the 'detection' of streaks in random sequences is attributed to a general misconception of chance" (p. 295), and this applies even to short sequences of random events. Individuals may even be stubbornly confident about these perceptions given the tendency to consider only confirmatory evidence when evaluating one's preexisting perceptions (Griffin & Tversky, 1992; Rabin & Schrag, 1999). This is likely to be especially true in the case of sequence interpretation,

since individuals update their beliefs incrementally based on new information, and may reinforce their prior predictions in future expectations (e. g., Hogarth & Einhorn, 1992).

These conflicting perspectives suggests that while interval *should* matter, there is little reason to expect one or the other—gradual intervals or rapid intervals between performance events—to more demonstrably shape perceived quality impressions in the minds of stakeholders.

3. Methods overview

In order to examine the performance characteristics of valence, order, and interval, our research design consists of both experimental and archival approaches. This mixed-methods approach allows us to test these characteristics in isolation as well as in combination with one another. Our first study takes an experimental approach to explore the influence of "order" and "valence" while controlling for interval. Our second study draws on archival data to replicate these findings around "order" while controlling for both valence (through an empirical measure) and interval (through study design). Finally, our third study employs an experimental approach to control for valence and further examine how "interval" shapes the "order" effect on perceived quality impressions. In study 3, we also test whether these order and interval effects persist over time.

4. Study 1 (Experiment 1 of 2)

To better isolate the effects of valence and order, our first study examines the interplay between valence and order while holding interval constant.

4.1. Sample

Our initial study begins with a controlled randomized experiment of 577 undergraduate business participants from a large Midwestern business school who were given course credit for participating in

management research studies. To prevent coercion, we employed the standard practice of providing an alternative, "opt out" assignment for those wishing not to participate in the study, and then dropped the one respondent who chose this from our sample.

4.2. Randomization conditions

We told participants they would be shown a vignette depicting product quality information about several products (since product quality information is arguably one of the most straightforward influencers of perceived quality), and then we randomly assigned participants to one of the nine independent variable conditions described below. In each condition, respondents read about a fictionalized Dutch manufacturer of HDTVs, "Voyark, Inc.", which specialized in "market-leading LED technology". They were told that over the past three years, Voyark had released three separate HDTV models, and they were then shown one of nine scenarios depicting a sequence of three explicit summaries of those product qualities with contrived "star ratings", quantities of consumers rating the product, and product names for each of the three products. These nine scenarios are depicted in **Fig. 1a**. Once they were exposed to one of these nine independent variable conditions, in order to capture our **dependent variable** of "perceived quality", we asked respondents to answer a 7-point Likert scale item: *"Based on these product quality ratings over the past 3 years, how would you rate Voyark's capability to develop high quality products?"* Response options ranged from 1, "very weak," to 7, "very strong". We crafted this measure to align with our definition of perceived quality as "the degree to which stakeholders positively evaluate an organization for the quality of its offerings, operations, or achievements", as well as the view that perceived quality is a consumer's subjective evaluation of a firm which can be gleaned from cumulative perceptions about the firm's performance vis-à-vis its offerings, operations, or achievements (e.g., Benjamin & Podolny, 1999; Parker et al., 2017; Rao, 1994; Rhee & Haunschild, 2006).[1]

1 But despite its merits in terms of parsimony in tapping the perceived quality of the firm, the limitations of this single-item measure are discussed, at length, in the discussion section.

	Worsening Scenarios	Improving Scenarios	Oscillating / Flat Scenarios
High Valence Scenarios	"Scenario 331" On December 11, **2014**, Voyark released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows: ★★★ (3 of 3 stars, from a total of 14,375 ratings) The following year, on December 11, **2015**, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows: ★★★ (3 of 3 stars, from a total of 15,563 ratings) Finally, on December 11, **2016**, Voyark released its XZ4, the latest model in its line of 48" HDTVs. The overall rating for this product was: ★☆☆ (1 of 3 stars, from a total of 17,632 ratings)	"Scenario 133" On December 11, **2014**, Voyark released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows: ★☆☆ (1 of 3 stars, from a total of 14,375 ratings) The following year, on December 11, **2015**, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows: ★★★ (3 of 3 stars, from a total of 15,563 ratings) Finally, on December 11, **2016**, Voyark released its XZ4, the latest model in its line of 48" HDTVs. The overall rating for this product was: ★★★ (3 of 3 stars, from a total of 17,632 ratings)	"Scenario 313" On December 11, **2014**, Voyark released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows: ★★★ (3 of 3 stars, from a total of 14,375 ratings) The following year, on December 11, **2015**, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows: ★☆☆ (1 of 3 stars, from a total of 15,563 ratings) Finally, on December 11, **2016**, Voyark released its XZ4, the latest model in its line of 48" HDTVs. The overall rating for this product was: ★★★ (3 of 3 stars, from a total of 17,632 ratings)
Moderate Valence Scenarios	"Scenario 321" On December 11, **2014**, Voyark released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows: ★★★ (3 of 3 stars, from a total of 14,375 ratings) The following year, on December 11, **2015**, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows: ★★☆ (2 of 3 stars, from a total of 15,563 ratings) Finally, on December 11, **2016**, Voyark released its XZ4, the latest model in its line of 48" HDTVs. The overall rating for this product was: ★☆☆ (1 of 3 stars, from a total of 17,632 ratings)	"Scenario 123" On December 11, **2014**, Voyark released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows: ★☆☆ (1 of 3 stars, from a total of 14,375 ratings) The following year, on December 11, **2015**, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows: ★★☆ (2 of 3 stars, from a total of 15,563 ratings) Finally, on December 11, **2016**, Voyark released its XZ4, the latest model in its line of 48" HDTVs. The overall rating for this product was: ★★★ (3 of 3 stars, from a total of 17,632 ratings)	"Scenario 222" On December 11, **2014**, Voyark released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows: ★★☆ (2 of 3 stars, from a total of 14,375 ratings) The following year, on December 11, **2015**, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows: ★★☆ (2 of 3 stars, from a total of 15,563 ratings) Finally, on December 11, **2016**, Voyark released its XZ4, the latest model in its line of 48" HDTVs. The overall rating for this product was: ★★☆ (2 of 3 stars, from a total of 17,632 ratings)
Low Valence Scenarios	"Scenario 311" On December 11, **2014**, Voyark released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows: ★★★ (3 of 3 stars, from a total of 14,375 ratings) The following year, on December 11, **2015**, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows: ★☆☆ (1 of 3 stars, from a total of 15,563 ratings) Finally, on December 11, **2016**, Voyark released its XZ4, the latest model in its line of 48" HDTVs. The overall rating for this product was: ★☆☆ (1 of 3 stars, from a total of 17,632 ratings)	"Scenario 113" On December 11, **2014**, Voyark released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows: ★☆☆ (1 of 3 stars, from a total of 14,375 ratings) The following year, on December 11, **2015**, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows: ★☆☆ (1 of 3 stars, from a total of 15,563 ratings) Finally, on December 11, **2016**, Voyark released its XZ4, the latest model in its line of 48" HDTVs. The overall rating for this product was: ★★★ (3 of 3 stars, from a total of 17,632 ratings)	"Scenario 131" On December 11, **2014**, Voyark released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows: ★☆☆ (1 of 3 stars, from a total of 14,375 ratings) The following year, on December 11, **2015**, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows: ★★★ (3 of 3 stars, from a total of 15,563 ratings) Finally, on December 11, **2016**, Voyark released its XZ4, the latest model in its line of 48" HDTVs. The overall rating for this product was: ★☆☆ (1 of 3 stars, from a total of 17,632 ratings)

Fig. 1a. Study 1: The Nine Product Quality Scenarios (Experiment 1 of 2).

In a three-event product quality sequence with three levels of quality ranging from "1" (poor quality), "2" (moderate quality), to "3" (high quality), there are potentially 27 distinct sequences containing three quality events. Logistical constraints limited us to the nine sequences depicted in Fig. 1a. After careful consideration, we initially selected three "overall high quality" conditions whose product qualities summed to a total of 7 (e.g., 3 + 3 + 1), three "overall modest quality" conditions which summed to a total of 6 (e.g., 3 + 2 + 1), and three "overall lower quality" conditions which summed to a total of 5 (e.g., 3 + 1 + 1).

Hereafter, these sequences are referenced as series of numbers that indicate the order and favorability of the quality events with which the respondents were presented. For instance, a scenario of 331, depicted in the figure above, denotes a scenario in which the respondent observed a 3-star product quality rating, followed by a 3-star product quality rating, followed by a 1-star product quality rating. Since all scenarios were "snapshots," respondents were not incrementally exposed to more information "over time", but they instead saw all three product quality events at once. This has the advantage of ruling out the influence of temporal recency bias due to memory effects, but perhaps it constitutes a tradeoff in realism (an issue we address in the third study, i.e., our second experiment). We test all models in Stata 14.2 using a one-way ANOVA with a post-estimation Tukey adjustment for multiple comparisons.

4.3. Findings

There are nine distinct scenarios, and the results of these are presented both graphically, in **Fig. 1b**, and in table form, in **Table 1**. Again, for the sake of readability, we have written these scenarios out here as number sequences, such that—for example—a scenario beginning with a product earning a quality rating "1 star" and followed by one with "3 stars" of product quality and then another product whose quality is "3 stars" is termed "scenario 133".

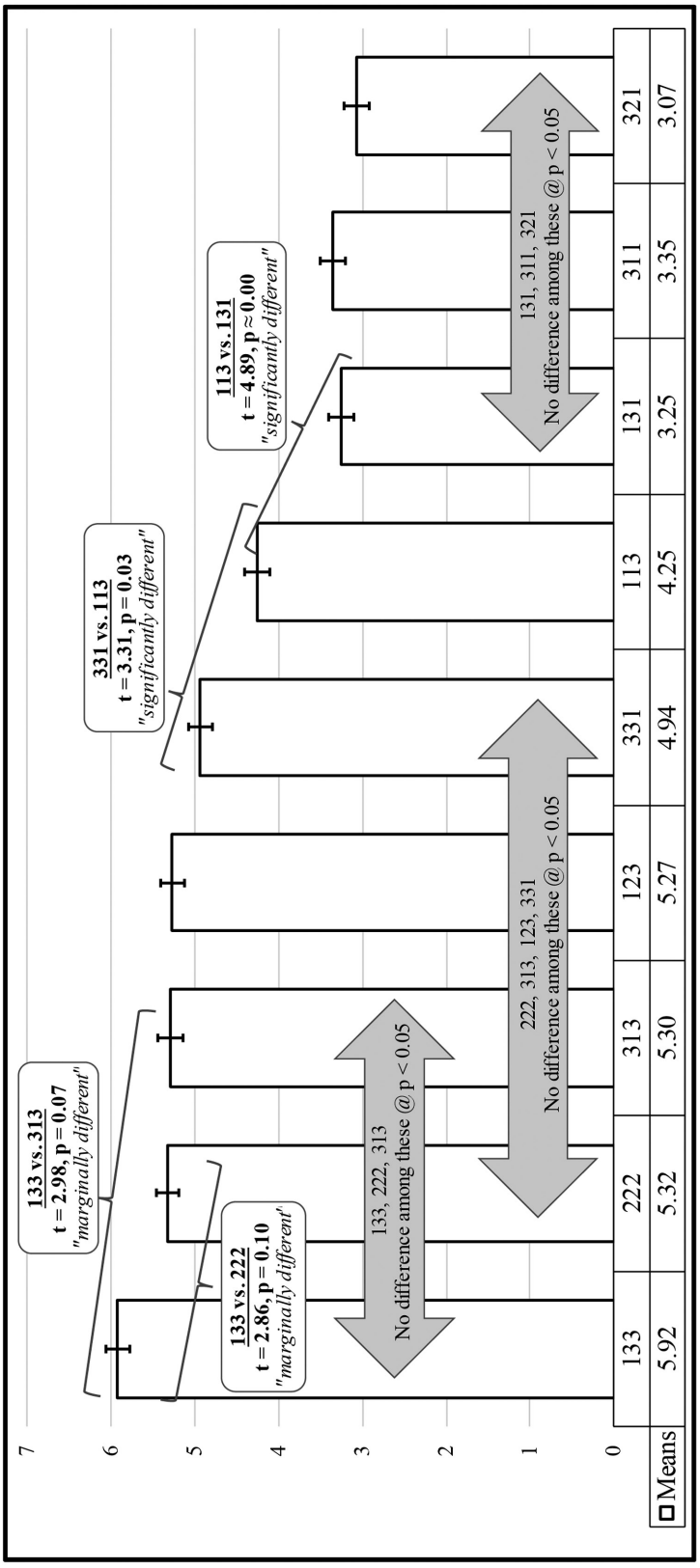

Fig. 1b. Results of Scenarios from Study 1 (Experiment 1 of 2). Number of observations = 577.

Table 1 Study 1: Summary statistics – means of outcome variables.

	Valence	*Trajectory*	*Means (Std. Err.)*	*Tukey Group(s)*
Scenario 133	High (7)	Improving	5.92 (0.15)	A
Scenario 222	Moderate (6)	*other*	5.32 (0.14)	AB
Scenario 313	High (7)	*other*	5.30 (0.15)	AB
Scenario 123	Moderate (6)	Improving	5.27 (0.14)	B
Scenario 331	High (7)	Worsening	4.94 (0.14)	B
Scenario 113	Low (5)	Improving	4.26 (0.15)	
Scenario 131	Low (5)	*other*	3.25 (0.14)	C
Scenario 311	Low (5)	Worsening	3.35 (0.15)	C
Scenario 321	Moderate (6)	Worsening	3.07 (0.15)	C

Number of observations = 577.

4.3.1. Wholly dominated sequences

One might expect scenarios that are superior in *both* valence *and* directional order to have better perceived quality than those that are inferior in both, and our results generally bore this out. There are two scenarios that are nominally superior to multiple other scenarios in both respects, and these are 133 (superior to 321, 131, 311, and 222) and 123 (superior to 131 and 311). Given that scenario 133 has a mean score of 4.92, this scenario scores higher perceived quality than 321 (mean = 2.07; t = 13.61, p = 0.00), higher than 131 (mean = 2.25; t = 12.87, p = 0.00), higher than 311 (mean = 2.35; t = 12.33, p = 0.00), but 133 *does not* score significantly higher perceived quality than 222 (mean = 4.32; t = 2.86; p = 0.101). Given that scenario 123 has a mean score of 4.27, this scenario scores higher perceived quality than 131 (mean = 2.25; t = 9.88, p = 0.00) and higher than 311 (mean = 2.35, t = 9.35, p = 0.00).

4.3.2. Valence matters, when controlling for order.

As we might have expected from prior research on perceived quality, we find that higher valence always outperforms lower valence when we hold order constant and examine valence alone. First, scenarios 113, 123, and 133 represent the uniformly improving sequences, and the higher valence scenario 133 outperforms 123 (means = 4.92 versus 4.27; t = 3.16, p = 0.04) and outperforms 113 (means = 4.92 versus 3.26; t = 7.99, p = 0.00). Second, scenarios 311, 321, and 331 represent

the uniformly worsening sequences, and the higher valence scenario 331 outperforms scenario 321 (means = 3.94 versus 2.07; t = 9.04, p = 0.00) and outperforms scenario 311 (means = 3.94 versus 2.35; t = 7.71, p = 0.00).

4.3.3. Effect of order holding valence constant

We find that improving scenarios always outperform worsening scenarios when we hold valence constant, but that oscillating and stable scenarios are often no different than same-valence improving / worsening scenarios.

To the first point, scenarios 133, 313, and 331 represent the higher valence sequences (each summing to 7), and among these, the improving sequence 133 outperforms 331 (means = 4.92 versus 3.94; t = 4.74, p = 0.00) and outperforms 313 (means = 4.92 versus 4.30; t = 2.98, p = 0.07). Scenarios 123, 222, and 321 represent the moderate valence sequences (each summing to 6), and among these, the improving sequence 123 outperforms 321 (means = 4.27 versus 2.07; t = 10.67, p = 0.00) and worsening sequence 321 underperforms stable sequence 222 (means = 2.07 versus 4.32; t = −10.86, p = 0.00). Scenarios 113, 131, and 311 represent the lower valence sequences (each summing to 5), and among these, the improving sequence 113 outperforms both 131 (means = 3.26 versus 2.25; t = 4.89, p = 0.00) and 311 (means = 3.26 versus 2.35; t = 4.38, p = 0.00).

The story is less clear when comparing these extreme scenarios to the more ambivalent stable or oscillating scenarios. Among the high valence scenarios, the worsening sequence 331 *does not* underperform the oscillating 313 (means = 3.94 versus 4.30; t = −1.74, p = 0.72), among moderate valence scenarios, sequence 123 does not outperform the stable scenario 222 (means = 4.27 versus 4.32; t = −0.28, p = 1.00), and among low valence scenarios, the worsening sequence 311 is no worse than the oscillating sequence 131 (means = 2.35 versus 2.25; t = 0.49, p = 1.00).

As such, our findings indicate that the improving sequences always outperform worsening sequences when valence is held constant, but that improving sequences do not necessarily outperform non-worsening sequences (e.g., 222), and worsening sequences do not necessarily underperform non-improving sequences (e.g., 313 or 131).

4.3.4. Where order overwhelms valence.

To this point, our "wholly dominated sequences" findings have indicated that valence tends to overrule differences in order, and our other findings also indicate that order does little to overcome differences in valence (i.e., improving scenarios often underperform worsening but higher valence scenarios, and worsening scenarios often outperform improving but lower valence scenarios). In fact, even the 133 scenario, which is both better than 222 in valence and represents an improvement over time, did not outperform 222 in this study. Thus, the majority of evidence from this study is that valence trumps order and that sometimes a superior order *and* a superior valence does not outperform a scenario that suggests *stability over time*.

However, the fact that scenario 113 does outperform the worsening scenario of 321 (mean = 3.26 versus 2.07; t = 5.74, p = 0.00) indicates that sometimes a scenario with lower valence but improving order can outperform a higher valence scenario.

4.4. Takeaways from study 1

We find that a strong positive signal sent by a uniformly improving scenario may sometimes outweigh the (perhaps weaker) negative signal sent by a poorer overall valence, at least in terms of perceived quality. A gradual march toward quality improvement might be weighted more heavily by respondents than a sequence that depicts generally better quality that appears to be "headed in the wrong direction".

5. Study 2 (archival study)

In the second study, we focus on the effect of order alone, holding the other two factors of valence and interval relatively constant. To do this, we employed an archival approach for several reasons. First, it gives us greater external validity to complement Study 1's advantages vis-`a-vis internal validity, and allows us to proxy for perceived quality performance indicators that could be seen to influence stakeholders' impressions. Second, the use of firm-years held interval constant,

and we are able to control for valence. In other words, we are able to isolate the effect of order in a way that optimizes external validity.

5.1. Sample and focal performance events

Our archival sample consists of all firms listed on the Boston Consulting Group (BCG) innovation rankings for the years 2006–2018. This fits with our definition of performance events which should influence the "perceived quality" of the firm, though we discuss the limitations of this approach in the Discussion section. We define perceived quality as the degree to which stakeholders positively evaluate an organization for the quality of its offerings, operations, or achievements. As we noted before, perceived quality pertains not just to product quality but to "being good" at what the firm purports to do (e.g., Rindova et al., 2005; Lange et al., 2011). When a firm becomes "known for innovation", for instance, innovation becomes a dimension of performance that stakeholders will come to expect, and on which stakeholders can assess the firm and hold it accountable. It is an important measure of the effectiveness of the firm's operations and achievements (and likely the merits of its offerings, as well). Therefore, we use BCG innovation rankings as the relevant performance events which we expect to influence perceived quality impressions, because this is both a critical component of the firm's "offerings or operations" and it is a metric that is evaluated across numerous firms. Again, we detail the merits and limitations of this further in the discussion section.

We have three approaches to measuring our predictor variable of order; as such, our sample size differed for each approach. The sample for the first and second approaches includes 136 firm-years across 34 firms, while the third approach includes 7044 firm-years across 737 firms. This includes an attenuation of observations due to our use of time lags as controls and firms dropping out of the BCG rankings after one year. We merged this with financial data from the Compustat North America fundamentals dataset.

5.2. Market impressions DV

We sought a measure that could directly tap perceived quality impressions, but there was no archival measure to which we had access that

(a) specifically and exclusively represented perceived quality and (b) was comparable across a sufficient number of firms in our data to prevent the reduction of statistical power. Thus, we chose to use Tobin's Q to proxy for the market's overall impression of the firm. We measure Tobin's Q following Chung and Pruitt (1994)—i.e., we sum market value, long term debt, and current liabilities, and divide the sum by total assets. Tobin's Q reflects the firm's value according to critical stakeholders, i.e., equity shareholders, and what the firm might make of its current financial situation given its past performance.

Tobin's Q is admittedly an imperfect measure of "perceived quality", alone, and we discuss this at length in the Discussion section. However, Tobin's Q is an appropriate proxy for the market's assessment of the firm's future capability. Furthermore, Tobin's Q should be influenced by the innovation ranking perturbations that serve as our focal performance events, and it is distinct from these innovation performance events. Moreover, there is some precedent for using Tobin's Q as a proxy for various social judgments in the finance literature (e.g., Cummins, Wei, & Xie, 2007; Gatzert, Schmit, & Kolb, 2016).

Moreover, for firms that are already represented in innovation rankings such as the BCG ranking, Tobin's Q is a relevant measure of perceived quality—with respect to innovation capability—because shareholders are likely to confer a higher level of importance on the firm's innovation ability when they evaluate how the firm might perform in the future. As such, despite its disadvantages, we can reasonably conclude that Tobin's Q is at least useful as a "market impressions" measures, and a one-step-removed proxy of perceived quality, and is therefore valuable when considered in concert with our other two experimental studies.

5.3. Order predictors

Our intent was to examine the effect of improving versus worsening (versus flat) orders on our market impressions proxy, Tobin's Q. For the sake of comprehensiveness, we took three approaches to operationalizing order changes. In all three cases, we examined how changes in rank from year t – 2 to t – 1 influenced our market impressions proxy — Tobin's Q — in year t. While study 1 employed three observation periods, our use of three periods in this archival study

attenuated our sample dramatically, causing us to default to a more straightforward, two-period window.

In the first and second approaches, our sample included only those firms that were ranked in both years t – 2 and t – 1, so these two approaches essentially examined the effect of perturbations *within* the BCG ranking list on Tobin's Q. The sample for the first and second approaches includes 136 firm-year observations for 34 firms. It excludes years 2012 and 2013. That is, BCG did not publish an innovation ranking list in 2011, which forced us to exclude the years 2012 and 2013 from our data due to a lack of consecutive years to create performance patterns.

The third approach included all firms and whether they were ranked in one or both years (i.e., a set of dummy scenarios), for a much larger sample. The sample for this approach includes 8180 firm-year observations for 740 firms across all years in the data. This also excludes years 2012 and 2013, for the same reason as above—no BCG ranking list was published in 2011.

5.3.1. First order approach: Continuous change in rank.

For our first approach, we computed the continuous change in rank from year t – 2 to year t – 1. For instance, a firm ranked #23 in t – 2 but ranked #17 in t – 1 would have a value of "6", whereas a firm that dropped from rank #17 in t – 2 to #23 in t – 1 would have a value of "□ 6". This is the most granular of the measurement approaches we took, but it also treats small and large changes in a continuous fashion, despite the fact that one might expect small changes to be essentially unnoticed by the market. For this reason, we also took our second and third approaches, detailed next.

5.3.2. Second order approach: rank-change thresholds

For our rank-change approach, for every firm year in our data, we examined the prior two firm-years and categorized as "improving order" those firms with an improvement in BCG ranking—by a certain threshold—from years t – 2 to t – 1. Similarly, we categorized as "worsening order" those firms with a worse BCG rank in year t – 1 compared to year t – 2, also by a certain threshold. We used numerous

rank-change thresholds to determine how big of a change was large enough to be "noticed", and where the effects seemed to be largest on Tobin's Q, our perceived quality proxy DV. We categorized as "no change" those changes that did not meet our specified thresholds for each of these models.

The thresholds we used in our separate models for this first approach included changes by at least 3, 4, 5, 6, and 7 ranks; we conjectured that a 5-rank change would be a reasonably noticeable change from a market perspective, and therefore set the thresholds centering on a 5-rank change. A change of 3 ranks is the least stringent, whereas a change of 7 ranks is the most stringent. For instance, a change from rank position #23 in year t – 2 to position #17 in year t – 1 amounts to an improvement of 6 rank positions. Similarly, a drop from rank #17 in year t – 2 to rank #23 in year t – 1 amounts to a worsening of 6 rank positions. The former would be picked up as "improvement" and the latter as "worsening" by our "at least 3, 4, 5, and 6-rank change" models, but categorized as "no change" by the stricter "at least 7-rank change" model.

5.3.3. Third order approach: Change in presence/absence on the ranking list

In our third approach, we dispensed with the granular differences in rank and examined the effect on Tobin's Q of whether, from year t – 2 to year t – 1, the firm: (a) became ranked, (b) became unranked, (c) remained ranked, or (d) remained unranked on the top 50 BCG ranking.

5.4. Control variables

All of our empirical models included year dummies (for year t), as well as several year t[2] financial controls: logged ***R&D spending***, logged ***revenues***, logged ***total employees***, logged ***total assets***,[3] and standardized ***net income***. We did not log net income because some values were

2 We included year t versions of these controls because of their expected influence on Tobin's Q in year t. However, when we instead use lagged versions of these controls (t – 1), there is no material difference in the results.

3 Although our Tobin's Q dependent variable includes total assets in the denominator, the results do not materially change when we drop this total assets control from our models.

negative and because it has a more normal distribution than that of the other financial controls. We orthogonalized the logged controls to address observed collinearity in our initial models.[4] For the first and second approaches, we also included ***prior year BCG rank*** (t – 1). Since this control limits the sample to only those firms that were ranked in at least year t – 1, we dropped this constraint for the third approach, since the third approach included many firms that became unranked in year t – 1, in order to examine the whole sample.

5.5. Analyses and results

Tables 2a-2c show descriptive statistics and bivariate correlations for the three approaches we employed to operationalize the "order" predictor.

We employed fixed effects panel regression in Stata 14.2, with robust standard errors clustered by firm. The results of these regressions are presented in Table 3. As discussed previously, we operationalized our improvement/worsening order predictor using three approaches, and the collective results from these three approaches paint a consistent picture of the influence of innovation rank-changes on market impressions (i.e., Tobin's Q).

5.5.1. First order approach: Continuous change in rank

As is indicated in **Table 3**, for the continuous measure, we found a positive and significant coefficient for "rank improvement" from t – 2 to t – 1, indicating that increases in year-on-year improvement in BCG ranking are associated with increases in Tobin's Q.

5.5.2. Second order approach: rank-change thresholds

For the rank-change predictors, we found significant effects in the improvement condition. Specifically, there was a significant ($p < 0.05$) coefficient for all threshold models except for the smallest, "at least 3-rank change" model, i.e., the change least likely to be noticed by the market did not generate a response in Tobin's Q. This is intuitive, and

4 Our models are also robust to dropping all but one size control.

Table 2a Study 2: Correlations and descriptive statistics for approach 1 (Continuous BCG Rank Improvement).

Approach 1	*Mean*	*SD*	*1*	*2*	*3*	*4*	*5*	*6*	*7*
1 Tobin's Q	2.50	1.48							
2 R&D Expense	4196.45	4075.34	0.10						
3 Revenues	89609.56	96974.25	–0.36	–0.66					
4 Employees	245.68	462.58	–0.39	–0.42	0.61				
5 Total Assets	99963.58	79860.38	–0.25	–0.09	0.13	–0.11			
6 Net Income	8910.74	10302.36	–0.05	0.18	0.30	–0.05	0.31		
7 Prior Year BCG Rank (t – 1)	18.46	13.01	–0.23	–0.24	0.09	0.11	0.06	–0.33	
8 BCG Rank Improvement (t – 2 to t – 1)	–0.32	10.77	0.10	0.17	–0.15	–0.14	0.08	0.05	–0.44

Correlations larger in magnitude than 0.23 are significant at $p < 0.01$ or better.

Descriptives are of untransformed variables, correlations are between variables entered in model.

Num. of observations = 136; Num. of firms = 34.

Table 2b Study 2: Correlations and Descriptive Statistics for Approach 2 (Changes by Rank Threshold).

Approach 2	*Mean*	*SD*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*10*	*12*	*14*	*16*
1 Tobin's Q	2.50	1.48												
2 R&D Expense	4196.45	4075.34	0.10											
3 Revenues	89609.56	96974.25	–0.36	–0.66										
4 Employees	245.68	462.58	–0.39	–0.42	0.61									
5 Total Assets	99963.58	79860.38	–0.25	–0.09	0.13	–0.11								
6 Net Income	8910.74	10302.36	–0.05	0.18	0.30	–0.05	0.31							
7 Prior Year BCG Rank (t – 1)	18.46	13.01	–0.23	–0.24	0.09	0.11	0.06	–0.33						
8 Up by 3 Ranks	0.32	0.47	0.00	–0.02	–0.06	–0.04	0.12	–0.12	–0.07					
9 Down by 3 Ranks	0.35	0.48	–0.08	–0.22	0.05	0.12	–0.06	–0.22	0.48	–0.50				
10 Up by 4 Ranks	0.28	0.45	–0.01	–0.06	–0.02	–0.02	0.07	–0.12	–0.09					
11 Down by 4 Ranks	0.31	0.46	–0.12	–0.24	0.07	0.06	–0.01	–0.20	0.50		–0.42			
12 Up by 5 Ranks	0.25	0.44	–0.00	–0.08	0.00	–0.06	0.07	–0.09	–0.07					
13 Down by 5 Ranks	0.28	0.45	–0.17	–0.26	0.10	0.09	0.03	–0.19	0.52			–0.36		
14 Up by 6 Ranks	0.21	0.41	0.02	–0.05	–0.04	–0.02	0.11	–0.09	–0.07					
15 Down by 6 Ranks	0.27	0.44	–0.15	–0.24	0.10	0.05	0.05	–0.18	0.52				–0.31	
16 Up by 7 Ranks	0.19	0.40	0.05	–0.08	–0.04	–0.05	0.11	–0.10	–0.08					
17 Down by 7 Ranks	0.24	0.43	–0.11	–0.25	0.13	0.09	0.06	–0.16	0.52					–0.28

Correlations larger in magnitude than 0.18 are significant at $p < 0.05$ or better. Correlations omitted where terms never coincide in a model together.
Descriptives are of untransformed variables, correlations are between variables entered in model.
Num. observations = 136; Num. firms = 34.

Table 2c Study 2: Correlations and Descriptive Statistics for Approach 3.

Approach 3	*Mean*	*SD*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
1 Tobin's Q	2.60	6.45									
2 R&D Expense	404.85	1232.27	–0.01								
3 Revenues	11228.11	31335.53	–0.18	–0.13							
4 Employees	35.02	105.34	0.06	–0.16	0.01						
5 Total Assets	12225.43	31446.51	–0.05	0.11	–0.13	0.16					
6 Net Income	824.00	3114.02	–0.02	0.26	0.29	0.04	0.11				
7 Unranked in Both t – 3 and t – 2	0.97	0.18	0.01	–0.25	–0.19	–0.08	–0.06	–0.44			
8 Ranked in Both t – 3 and t – 2	0.01	0.09	–0.00	0.19	0.15	0.07	0.03	0.36	–0.75		
9 Unranked in t – 3, Ranked in t – 2	0.01	0.08	–0.00	0.11	0.08	0.03	0.04	0.20	–0.50	–0.01	
10 Ranked in t – 3, Unranked in t – 2	0.02	0.14	–0.01	0.10	0.08	0.04	0.04	0.13	–0.42	–0.01	–0.01

Correlations larger in magnitude than 0.03 are significant at $p < 0.05$ or better.
Descriptives are of untransformed variables, correlations are between variables entered in model.
Num. observations = 7044; Num. firms = 737.

Table 3 Study 2: Fixed effects panel regression of Tobin's Q.

	Approach 1				Approach 2										Approach 3			
	Model 1A *Controls*		Model 1B *Full Model*		Model 2A *3 rank-changes*		Model 2B *4 rank-changes*		Model 2C *5 rank-changes*		Model 2D *6 rank-changes*		Model 2E *7 rank-changes*		Model 3A *Controls*		Model 3B *Full Model*	
	b/se	*p*	b/se	*p*	b/se	*p*	b/se	*p*	b/se	*p*	b/se	*p*	b/se	*p*	b/se	*p*	b/se	*p*
R&D Expense (natural log)	–4.23	0.05	–3.87	0.04	–3.95	0.04	–3.76	0.04	–3.64	0.04	–3.40	0.04	–3.81	0.03	0.85	0.38	0.86	0.38
	2.06		1.84		1.83		1.74		1.70		1.56		1.71		0.96		0.96	
Revenues (natural log)	–1.88	0.24	–1.80	0.23	–1.92	0.23	–2.21	0.19	–2.31	0.18	–2.42	0.15	–2.13	0.18	–3.73	0.13	–3.73	0.13
	1.57		1.46		1.59		1.66		1.69		1.62		1.56		2.45		2.44	
Employees (natural log)	1.86	0.05	1.76	0.05	1.77	0.04	1.73	0.04	1.75	0.04	1.72	0.03	1.83	0.03	1.13	0.32	1.13	0.32
	0.91		0.85		0.85		0.82		0.83		0.78		0.81		1.14		1.14	
Total Assets (natural log)	–1.27	0.00	–1.22	0.00	–1.30	0.00	–1.30	0.00	–1.26	0.00	–1.27	0.00	–1.33	0.00	–2.55	0.07	–2.55	0.07
	0.31		0.32		0.33		0.35		0.33		0.34		0.33		1.38		1.38	
Net Income (standardized)	0.23	0.14	0.21	0.14	0.22	0.13	0.22	0.13	0.21	0.13	0.20	0.13	0.21	0.12	0.05	0.34	0.05	0.32
	0.15		0.14		0.14		0.14		0.14		0.13		0.13		0.05		0.05	
Prior Year BCG Rank (t – 1)	0.01	0.39	0.02	0.05	0.01	0.13	0.02	0.07	0.02	0.06	0.02	0.06	0.01	0.18				
	0.01		0.01		0.01		0.01		0.01		0.01		0.01					
BCG Rank Improvement (t – 2 to t – 1)			**0.02**	**0.06**														
0.01																		
Up by X ranks (dummy)					0.12	0.34	**0.25**	**0.04**	**0.28**	**0.09**	**0.48**	**0.01**	**0.65**	**0.02**				
					0.13		0.12		0.16		0.18		0.25					
Down by X ranks (dummy)					–0.26	0.31	–0.27	0.29	–0.30	0.22	–0.25	0.37	0.12	0.60				
					0.25		0.26		0.24		0.28		0.22					
Ranked in Both (dummy)																	–0.17	0.62
																	0.34	
Unranked, then Ranked (dummy)																	–0.03	0.91
																	0.27	
Ranked, then Unranked (dummy)																	**–0.42**	**0.09**
																	0.25	
Constant	8.53	0.01	7.94	0.00	8.27	0.01	8.25	0.01	8.15	0.00	7.98	0.00	8.11	0.00	0.32	0.80	0.33	0.80
	2.90		2.58		2.74		2.71		2.65		2.49		2.51		1.28		1.28	
Model F	104.60	0.00	86.89	0.00	62.21	0.00	265.40	0.00	191.40	0.00	331.5	0.00	135.2	0.00	13.47	0.00	12.13	0.00
Num of Observations	136		136		136		136		136		136		136		7044		7044	
Num of Firms	34		34		34		34		34		34		34		737		737	

Robust standard errors listed below beta coefficients. Significant main effects bolded. Year dummies omitted for parsimony.

in fact, the improvement condition coefficients tended to get larger and the p-values smaller as the rank change thresholds became larger, with the strongest results for the 6- and 7-rank change threshold models. This suggests that larger or perhaps "more noticeable" changes in rank are more likely to influence the market's perceptions of the firm's capability vis-à-vis Tobin's Q. However, there was no significant ($p < 0.05$) coefficient for the worsening condition.

In post-hoc tests for the first two order approaches we conducted linear tests on the strength of the coefficients to determine whether valence (i.e., prior rank) or order (i.e., rank change) had a stronger predictive effect on Tobin's Q. Our test of the first approach's model yielded no significant difference between the two coefficients ($F = 2.31$, $p = 0.12$), but the post estimation F tests of the discrete linear effects within the second approach's models did indicate that the "increase by X ranks" version of the order predictor had stronger effects than "Prior Year BCG Rank (t – 1)" within several of the models. That is, in Model 2B, the "up by 4 ranks" coefficient was 0.25 ($p = 0.04$), the "Prior Year BCG Rank (t – 1)" coefficient was 0.02 ($p = 0.07$), and the "up by 4 ranks" coefficient was statistically stronger ($F = 4.22$, $p = 0.05$). Similarly, in Model 2C, the "up by 5 ranks" coefficient was 0.28 ($p = 0.09$), the "Prior Year BCG Rank (t – 1)" coefficient was 0.02 ($p = 0.06$), and the "up by 5 ranks" coefficient was statistically stronger, although this difference was marginally significant ($F = 2.91$, $p = 0.098$). In Model 2D, the "up by 6 ranks" coefficient was 0.48 ($p = 0.01$), the "Prior Year BCG Rank (t – 1)" coefficient was 0.02 ($p = 0.06$), and the "up by 6 ranks" coefficient was statistically stronger than the prior year coefficient ($F = 7.29$, $p = 0.01$). Finally, in Model 2E, the "up by 7 ranks" coefficient was 0.65 ($p = 0.02$), the "Prior Year BCG Rank (t – 1)" coefficient was 0.01 ($p = 0.18$), and the "up by 7 ranks" coefficient was statistically stronger ($F = 6.37$, $p = 0.017$). However, in Model 2A, the coefficient for "up by 3 ranks" was not significant, and neither was the "prior year BCG rank (t – 1)" coefficient.

5.5.3. Third order approach: Change in presence/absence on the ranking list

In our third approach, we found that there was no difference in Tobin's Q between being "unranked in both t – 2 and t – 1" (the base case) and

either "becoming ranked in t – 1" or "being ranked in both t – 2 and t – 1". However, there is a significant and negative effect of "becoming unranked in t – 1" after having been ranked in t – 2. When we reset the base case to "becoming unranked in t – 1", we see there is no difference between this scenario and "being ranked in both" or "becoming ranked in t – 1". Thus, it appears from our data that it is materially worse to be ranked-then-unranked than it is to have *never been ranked* in either of those two years, but no better or worse to lose one's ranking than it is to gain a ranking or to have always been ranked.

5.5.4. Synthesis of the order measurement approaches

All of these order measurement approaches indicate that (a) improvement tends to be more influential than worsening orders for firms that have *remained* ranked over time, but that (b) while becoming ranked seems to offer no benefit over having never been ranked, (c) dropping off the ranking is worse than having never been ranked.

5.6. Synthesis of studies 1 and 2 and rationale for study 3

We examined valence and order in study 1 and controlled for the influence of valence in study 2, and found evidence that order seems to have a contingent effect on perceptions. We therefore decided to conduct a third study in which we held valence constant while varying the order as well as the *interval* of performance events in various sequences, allowing us to test (a) the effect of interval holding order constant and (b) the effect of order holding interval constant.

6. Study 3 (experiment 2 of 2)

6.1. Sample

Our sample for Study 3 consisted of 471 undergraduate business participants from a large Midwestern business school who were given course credit for participating in management research studies. To prevent coercion, we employed the standard practice of providing an alternative, "opt out" assignment for those wishing not to participate in the study, but no respondent chose this option.

6.2. Randomization conditions

6.2.1. Vignette primer.

Participants were shown a vignette depicting product quality information about several products released by a fictionalized Swedish manufacturer of HDTVs, "Keltrek, Inc.", which specialized in "market-leading LED technology". Respondents were randomly assigned to one of the four scenarios consisting of five "years" each. Every "year" of information about the company's progress was shown on a separate screen in chronological order. For every "year" of this scenario, participants saw a paragraph or two about products released in the prior year, and how highly consumers had rated that year's product(s), from 1 star (minimum) to 4 stars (maximum). We controlled for valence by ensuring that each scenario had an identical total sum of "stars" across the five years of the scenario.

6.2.2. Perceived quality of the firm in each "year"

After each year, respondents were asked to answer a perceived quality question so that the effect of the information sequence on our **perceived quality dependent variable** could be assessed at every point in time. Similar to Study 1, respondents were asked a 7-point Likert scale item: *"…how would you rate Keltrek's capability to develop high quality products?"* Response options ranged from "very weak" (1) to "very strong" (7). The same advantages and caveats of this measure that we noted in Study 1 apply here, and we elaborate these later in the limitations section.

6.2.3. Four scenarios by order and interval

Years 2, 3, and 4 of the scenario were manipulated by two attributes: (a) **order**, whether the quality of successive products improved or worsened over time and (b) **interval**, whether the products were released gradually—i.e., one each year for years 2, 3, and 4—or rapidly— i.e., no products released in years 2 or 3 but all three products released four months apart (April, August, and December) in year 4, and shown to the respondents all at once at the end of that "year". All four scenarios are depicted in **Fig. 2a** and **Fig. 2b**.

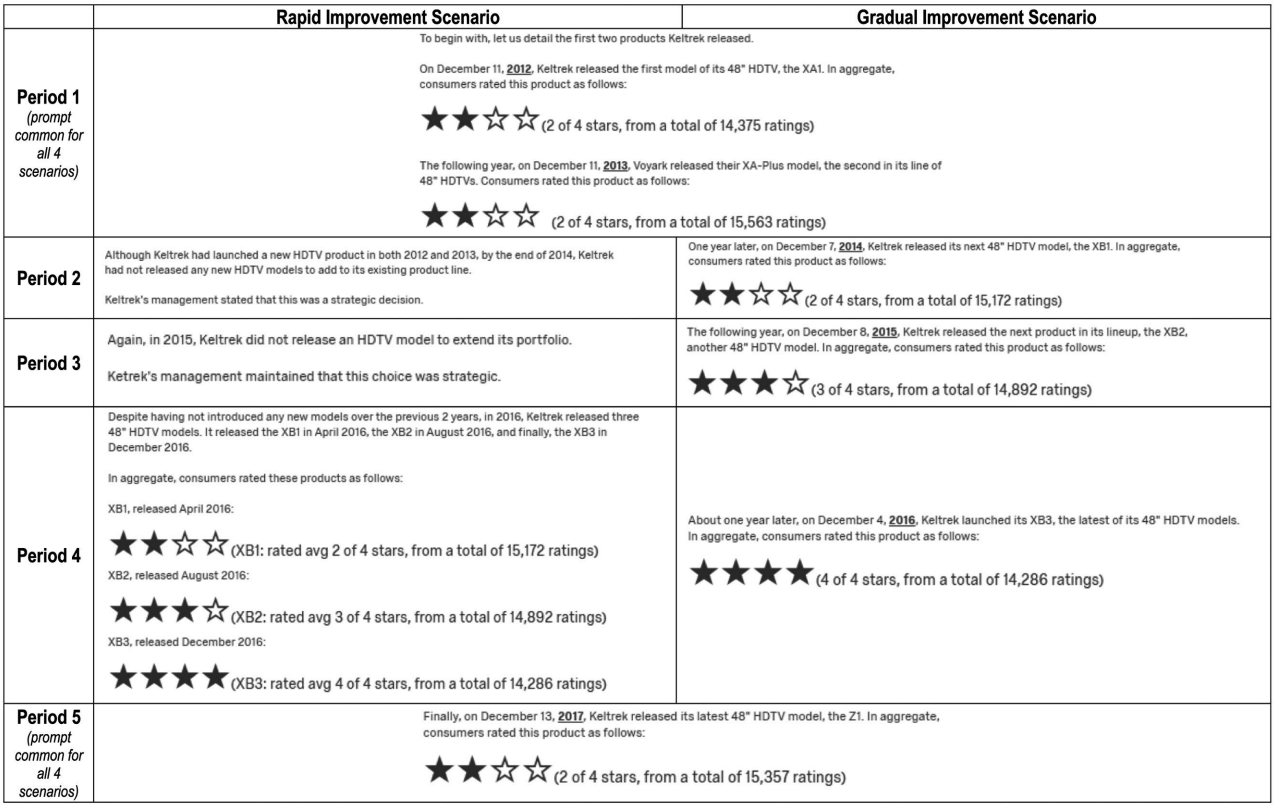

Fig. 2a. Study 3: "Improvement" Scenarios (Experiment 2 of 2).

6.2.4. Identical first and fifth "years"

To control for the potentially disproportionate influence of first observation, the first year was identical across all four scenarios. That is, for all scenarios, the first year a retrospective description of the two prior years, in which the company released two products—one each year—which were rated "2 stars" each.

We also wanted to determine the durability of the scenarios' effects on stakeholders' perceived quality of the firm. That is, we expect responses to differ after the year 4 mark, since this is the point at which respondents had been exposed to all the information distinguishing the four scenarios from one another. However, it is possible that these differences would *not* persist once the respondents had been exposed to new information which might "wipe clean" those prior effects. To test whether the scenario effects persisted on perceived quality, we added a fifth period—identical across all four scenarios—in which respondents saw a "2 star" product.

<table>
<tr><th></th><th>Rapid Worsening Scenario</th><th>Gradual Worsening Scenario</th></tr>
<tr><td>Period 1
(prompt common for all 4 scenarios)</td><td colspan="2">To begin with, let us detail the first two products Keltrek released.

On December 11, 2012, Keltrek released the first model of its 48" HDTV, the XA1. In aggregate, consumers rated this product as follows:
★★☆☆ (2 of 4 stars, from a total of 14,375 ratings)

The following year, on December 11, 2013, Voyark released their XA-Plus model, the second in its line of 48" HDTVs. Consumers rated this product as follows:
★★☆☆ (2 of 4 stars, from a total of 15,563 ratings)</td></tr>
<tr><td>Period 2</td><td>Although Keltrek had launched a new HDTV product in both 2012 and 2013, by the end of 2014, Keltrek had not released any new HDTV models to add to its existing product line.

Keltrek's management stated that this was a strategic decision.</td><td>One year later, on December 7, 2014, Keltrek released its next 48" HDTV model, the XB1. In aggregate, consumers rated this product as follows:
★★★★ (4 of 4 stars, from a total of 15,172 ratings)</td></tr>
<tr><td>Period 3</td><td>Again, in 2015, Keltrek did not release an HDTV model to extend its portfolio.

Ketrek's management maintained that this choice was strategic.</td><td>The following year, on December 8, 2015, Keltrek released the next product in its lineup, the XB2, another 48" HDTV model. In aggregate, consumers rated this product as follows:
★★★☆ (3 of 4 stars, from a total of 14,892 ratings)</td></tr>
<tr><td>Period 4</td><td>Despite having not introduced any new models over the previous 2 years, in 2016, Keltrek released three 48" HDTV models. It released the XB1 in April 2016, the XB2 in August 2016, and finally, the XB3 in December 2016.

In aggregate, consumers rated these products as follows:

XB1, released April 2016:
★★★★ (XB1: rated avg 4 of 4 stars, from a total of 15,172 ratings)
XB2, released August 2016:
★★★☆ (XB2: rated avg 3 of 4 stars, from a total of 14,892 ratings)
XB3, released December 2016:
★★☆☆ (XB3: rated avg 2 of 4 stars, from a total of 14,286 ratings)</td><td>About one year later, on December 4, 2016, Keltrek launched its XB3, the latest of its 48" HDTV models. In aggregate, consumers rated this product as follows:
★★☆☆ (2 of 4 stars, from a total of 14,286 ratings)</td></tr>
<tr><td>Period 5
(prompt common for all 4 scenarios)</td><td colspan="2">Finally, on December 13, 2017, Keltrek released its latest 48" HDTV model, the Z1. In aggregate, consumers rated this product as follows:
★★☆☆ (2 of 4 stars, from a total of 15,357 ratings)</td></tr>
</table>

Fig. 2b. Study 3: "Worsening" Scenarios (Experiment 2 of 2).

6.3. Analysis and results

We tested the effects of these conditions in Stata 14.2 using a one-way ANOVA with a post-estimation Tukey adjustment for multiple comparisons. We did separate tests at each year, to determine the differences in perceptions at each point in time throughout the five years. These results revealed several interesting findings, and the mean perceived quality scores at each point of observation are provided in the **Table 4** and illustrated graphically in **Fig. 2c**.

6.3.1. Effects of order

We were first interested in assessing the effect of performance order on perceived quality. For this reason, we focus on the scores after year 4, when participants had been exposed to all distinct information for the improving or worsening scenarios. After year 4, the

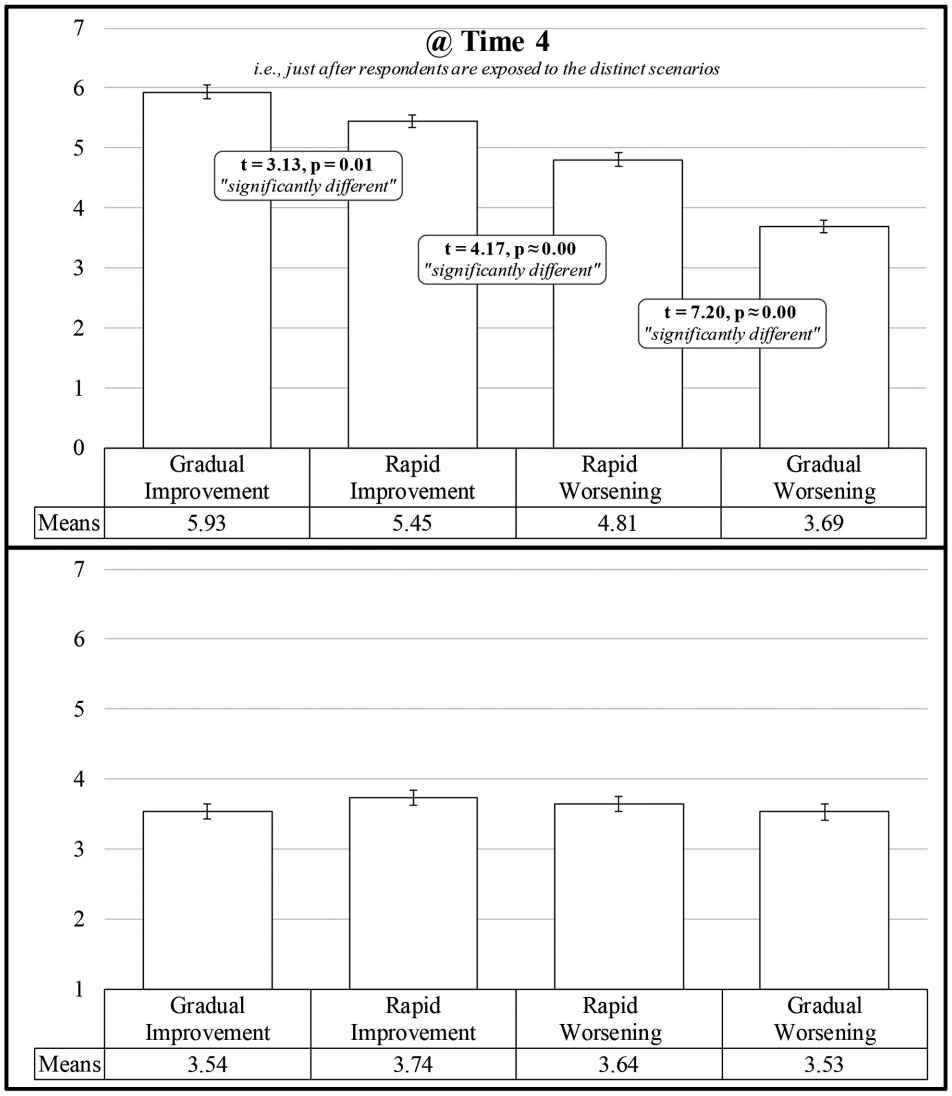

Fig. 2c. Study 3: Perceived Quality Scores by Condition (Experiment 2 of 2). *Number of observations = 471.*

perceived quality scores for the "gradual improvement" scenario were significantly greater than those of the "gradual worsening" scenario (t = 14.48, p = 0.000), and the scores for the "rapid improvement" scenario were also greater than those of the "rapid worsening" scenario (t = 4.17, p = 0.000), which essentially replicated our results from our archival study. That is, improving orders always seem to outperform worsening orders.

6.3.2. Effects of interval

To test the effect of interval, we examined the differences in effects for the same order. We found that the mean perceived quality scores for "gradually worsening" scenario (mean = 3.69) were significantly lower (t = 7.20, p = 0.000) than those of the "rapidly worsening" scenario (mean = 4.81). Conversely, the mean perceived quality score for the "gradually improving" scenario (mean = 5.93) were significantly higher (t = 3.13, p = 0.010) than those of the "rapidly improving" scenario (mean = 5.45). This suggests that when order is held constant, a gradual revelation of quality information yields a stronger effect on perceived quality than a rapid revelation.

6.3.3. No persistence of differences in year 5

Again, we added the fifth year to determine whether these effects would persist once respondents were shown new information that might override the scenario to which they had been exposed. Interestingly, our test of the effects after year 5—during which all scenarios showed the release of a "2 star" product—indicated *no significant differences* in perceived quality scores at this period. Since responses were cumulative, this suggests that the effect of the scenario revealed over years 2, 3, and 4 was wiped away by that fifth year of information.

7. Discussion

We examined the disparate and synergistic effects of valence, order, and interval of a sequence of performance events on stakeholders' perceived quality impressions about a firm. These investigations lead us to draw several key discoveries, which we first list and then elaborate in the following section.

7.1. Disappointment trumps complacency

One consistent finding is that a "fall from grace" (e.g., losing an existing ranking) may be worse than never having been ranked at all.

In study 2, our third order operationalization—i.e., whether the firm was ranked on the BCG innovation ranking the prior year, whether it had joined, fallen off, or remained unranked in both years—indicated that falling off the rankings seemed to have a more negative impact on perceived quality than never having been ranked at all. Notably, we controlled for valence in study 2 in the other two operationalizations of order by including the "prior rank" term in the model as a control variable. However, whether the firm is ranked—at all—is a nuance that cannot be both controlled for and tested in the same model, and our results indicate that such a large change (i.e., dropping out of the rankings) presents a unique hazard to stakeholders' perceived quality of the firm.

7.1.1. Liability of high perceived quality for performance

This particular finding is consistent with studies in the social evaluations literature on the "liability of good reputation" (e.g., Rhee & Haunschild, 2006; Zavyalova et al., 2012). Specifically, firms that present evidence of their capability in some area but then fail to perpetuate this over time may be worse off (in terms of stakeholder judgments) than if they did not set such high expectations.

7.1.2. Nuances of changes in performance on stakeholders' perceived quality of the firm

There is growing evidence of the importance of living up to stakeholder expectations, yet less is known about how specific performance incidents shape a firm's perceived quality in the eyes of stakeholders beyond valence alone (e.g., Rindova et al., 2005). Thus, there are two gaps in the literature: the first pertaining to the granularity of the circumstances studied, with much less attention to non-crisis situations which describe the experiences of most firms most of the time (Titus et al., 2018); and the second pertaining to how perceived quality is *formed* from these nuances in performance events. While scholars have looked at how performance effects may be attributable to deviations from a *reputational* standard (e.g., Zavyalova et al., 2012; Zavyalova et al., 2016), we have yet to examine how perceived quality is shaped by perturbations in performance, itself.

7.2. Valence and order both matter

The notion that individuals strive to "make meaning from random noise", even when no such meaning exists, suggests that firms with performance sequences that seem to indicate an increasing (decreasing) order of performance may benefit (suffer) according to the apparent order of these performances. For instance, if a firm with lower capability has a sequence of performances that steadily improve, the firm may be more favorably perceived by stakeholders than a higher-capability firm with a seemingly worsening order of performances. We find consistent evidence across all three studies that "order matters" for impressions of a firm's perceived quality. Study 1 demonstrated that order matters within same-valence comparisons in a controlled, internally valid setting. In study 2, we controlled for valence using an archival approach, and still found that order matters in an externally valid setting of market observers explicitly valuing the firm after observing the innovation rank changes of prominent firms—especially when those rank changes are larger, rather than smaller. Finally, our evidence from study 3 echoes these findings by replicating this order effect when examining just four separate scenarios. This is consistent with psychology literature on judgment and decision making, in which individuals are described as interpreting information which they perceive as meaningful in a heuristic fashion. It is also consistent with firm-level research on reputation, though our findings yield novel insight as well. We detail each of these research bases, in turn.

7.2.1. Ambiguity on the primacy of order versus valence on perceived quality

Although order explains incremental variance in perceived quality beyond valence alone, there is a limit to this order effect. Our empirical investigation produced nuanced findings. On the one hand, we find evidence that supports the primacy of "order over valence"—in our posthoc tests in Study 2, which showed that "rank change" almost always had a stronger predictive effect on Tobin's Q than prior rank, alone. However, we also find evidence that supports the primacy of "valence over order"—in our tests in Study 1, in which order beat

valence only once (scenario 113 versus 321) in head-to-head comparisons of order and valence on perceived quality.

7.2.2. Short-lived effects of order and interval

Perhaps most importantly, our results in study 3 revealed that the durability of the effects of order and interval is short-lived—when exposed to new information, respondents appear to forget about the disparate information from scenarios of disparate orders/intervals that led them to markedly different perceived quality impressions. We included this persistence check in the third study to attempt to capture the longevity of these effects, and future researchers in this vein should note the importance of determining the longevity of the perceived quality effects they purport to demonstrate. While we know a fair amount about how social judgments such as reputation influence performance, we know much less about how perceived quality is formed. There may be circumstances under which the effects of order and interval on stakeholders' perceived quality of the firm persist even in the face of conflicting information, but this discovery suggests that valence— and in particular, *recent* valence—is paramount for perceived quality formation.

7.2.3. Slow revelation trumps rapid revelation for perceived quality effects

Our results provide clear evidence that interval also matters for perceived quality formation. In study 3—where we hold valence constant— we find clear evidence that a steady and protracted interval of performance events yields a stronger effect of order. That is, rapidly improving orders underperform gradually improving orders. Similarly, gradually worsening orders underperform rapidly worsening orders. This is likely because stronger impressions are formed from the gradual revelation of performance evidence than more rapid revelation of such evidence.

Individuals are awash in noisy information environments and may sometimes miss signals if they occur too rapidly (e.g., Chaiken, 1980; Wyer Jr, 2003), and more gradual revelation may allow a sequence of performances to be incrementally confirmed in the mind of the

observer (Wyer, Adaval, & Colcombe, 2002). By contrast, if signals are received too rapidly—as in the case of a rapid revelation interval—such signals are either (a) unlikely to be noticed by the observer amid numerous other demands on their attention or (b) unlikely to have the same degree of influence on the observer's evolving impression about the entity in question (Chaiken, 1980; Chaiken, 1989; Chaiken & Ledgerwood, 2011).

At the firm level, performance sequences generally fall somewhere on this spectrum from "gradual revelation"—such as our condition of three product releases, once annually over three years—to "rapid revelation"— such as our condition of no performance events for two years and then three product releases in the next year. Firms may not always have control over how rapidly information is released, but they can change the pace of their product introduction (e.g., Parker et al., 2017), and this might have an effect on perceived quality formation.

In terms of the gradual versus rapid revelation of negative information, this is consistent with the notion of "scrutiny bundling" discussed in recent literature (Titus et al., 2018), in which firms sometimes hasten activities that can attract negative scrutiny when media coverage is already sour. When information is expected to generate a negative perceptual response in stakeholders, "ripping off the Band-Aid" at once may be more effective at alleviating a negative perceptual burden than gradually revealing negative information about the firm.

Conversely, our results indicate that the protracted revelation of improving news about quality has a stronger positive effect on perceptions of the firm's quality capability than would the rapid revelation of a positive trend in quality. There is a certain credibility in demonstrating a gradual increase in quality, over time, rather than exhibiting evidence of a skyrocketing quality without such a gradual revelation. Even in cases in which the information is revealed in the very same order and the quality endpoint is held constant, the gradual revelation of progressively improving product quality evidence will more positively shape stakeholder perceptions than a rapid revelation of a dramatic improvement in product quality evidence. Notably, this has yet to be examined at the organizational level.

8. Limitations and future research

As with any study, our work is subject to certain limitations. We explicitly detail many of these below so that readers are aware of the assumptions underpinning our analyses and discussions.

8.1. Tobin's Q as a proxy for market impressions stemming from perceived quality

The nuanced nature of our findings stems from the diverse contexts that we study, and some caution must be taken in interpreting our results. In several places, we made the "best fitting choice" regarding the conceptual elements we intended to study. For instance, in Study 2, we chose Tobin's Q to proxy for market impressions, which we characterized as one-step-removed from perceived quality. Tobin's Q could be used to measures numerous concepts and latent constructs, and is not exclusively a proxy for "perceived quality". For this reason, we do not claim to directly measure perceived quality, but rather the market impressions we expected to be shaped by perceived quality.

However, Study 2 revealed some interesting findings about the effect of order that—when taken in concert with Study 1 and Study 3—are informative despite the limitations of the measurement approach we were forced to use in this archival context. We should also note that in this respect, our use of Tobin's Q to proxy for the effect of perceived quality on market impressions is similar to using R&D investment to operationalize myriad constructs across the organization sciences (Bromiley, Rau, & Zhang, 2017; Ketchen Jr., Ireland, & Baker, 2013). Archival contexts are inherently limited in their ability to precisely tap the concepts of interest, but they allow for greater external validity. We also believe that our experimental studies helped complement these disadvantages due to their greater internal validity.

8.2. Single item measure of perceived quality

In our two experiments, our measures of perceived quality were single-item measures rather than multiple item measures that some researchers use for well-established, latent constructs. Our results should be received and interpreted in light of this limitation. In

general, we agree that researchers should use multiple item scales, particularly when those multiple items have been evaluated for their factor loadings in a careful, methodical, construct validation process (e.g., MacKenzie, Podsakoff, & Podsakoff, 2011). Past work on perceived quality at the firm level has mainly used perceptual ratings by stakeholder-respondents, usually averaged together, to proxy for perceived quality (e.g., Rindova et al., 2005; Rhee & Haunschild, 2006; Parker et al., 2017).

In our single-item perceived quality measures for both Study 1 and Study 3, we chose to use a psychometric item with linkages both to these prior operationalizations and to the conceptual definition of perceived quality. We did so because of the parsimony afforded by such an approach (e.g., Hoeppner, Kelly, Urbanoski, & Slaymaker, 2011), particularly since there is little agreement as to the conceptual domain of firm-level "perceived quality" from a psychometric standpoint. In fact, there is little consensus about the psychometric properties of many social judgment constructs at the firm level and the items that represent their content domains, although conceptual efforts have gained momentum recently (e.g., Ravasi et al., 2018) and psychometric work has recently begun to address this issue (e.g., Bitektine, Hill, Song, & Vandenberghe, 2020). Moreover, since we conducted multiple studies with some consistency among the results, we are less concerned about the impact of this single-item measure than we would be if we had only conducted a single study with a single-item measure to capture perceived quality.

Nevertheless, future research should, where possible, employ multiple item measures for perceived quality—especially if only one study is being conducted—so that the veracity of the results is more defensible.

8.3. Snapshot versus experience over time

In both of the experimental studies, our scenarios were generally retrospective snapshots of how a firm performed in terms of its product quality. In Study 1, the information was presented "all at once", so that the respondent saw a scenario only at one point, and was not taken through its evolution step by step. In Study 3, the information was gradually revealed as the story of the firm's experience was told year by year. However, in both of these cases, the performance was

characterized as having occurred in the past. Readers should note that this is not the same as an experimental design which actually incorporates real-time firm performances and the respondents' impressions resulting from those performances. Our approach was, nevertheless, useful to understand how a firm's perceived quality is shaped by various patterns of retrospective information about its performance. In reality, stakeholders may not fully experience or observe a series of events that demonstrate a firm's quality. However, they can readily observe the retrospective snapshot of the firm's past performance pattern at any point in time. It is this retrospective snapshot that reveals a firm's existing capability of creating value for stakeholders, by which the market evaluates/anticipates the firm's future capacity. So, compared to the step-by-step real-time experience of performance events, a retrospective snapshot probably matters more in shaping stakeholders' perceived quality of a firm, and thus, more practically important. Moreover, a "step by step" experience approach is more likely to be dominated by recency effects (Hogarth & Einhorn, 1992). Since respondents may not be able to remember all the past quality signals clearly, their perception is most likely to show recency effects. Presenting a retrospective snapshot of past quality signals to respondents can avoid this potential bias. Future research could extend our work by crafting a design which incorporates real-world revelation of performance events over a longer period, although this is likely to be far more costly to create.[5]

9. Conclusion

Our study is an exploratory analysis of one facet of the landscape of perception formation, examining how a firm's "perceived quality" takes shape. We address whether and how much the order, interval, and valence of a firm's performance events shape its perceived quality in the minds of stakeholders. Due to the lack of theoretical clarity on the topics we address, especially at the organizational level of analysis, our exploratory analysis offers unique insight into the perceived quality formation process that is not possible via traditional

5 We appreciate this insightful comment from an anonymous reviewer.

hypothesis testing methods. In a time in which public perceptions of a company's capabilities are paramount, managers are increasingly benefited by tips that might advantage these perceptions, especially when all that is required is a shifting of timing in activities in which they are already engaged. We aimed to make a meaningful first effort in helping shed light on this important set of phenomena. We hope that our work sparks future attention and interest among scholars interested in this stream of work around perception formation—and perceived quality formation, in particular—and that our insights (and missteps) help advance the quality of subsequent research efforts in this vein and related areas of research.

.....

Acknowledgments We would like to thank Angela Garbacz, Nate Gingery, and Kevin Shinn for sharing their insights during the initial stages of this research project.

References

Basdeo, D. K., Smith, K. G., Grimm, C. M., Rindova, V. P., & Derfus, P. J. (2006). The impact of market actions on firm reputation. *Strategic Management Journal, 27*(12), 1205–1219.

Benjamin, B. A., & Podolny, J. M. (1999). Status, quality, and social order in the California wine industry. *Administrative Science Quarterly, 44*(3), 563–589.

Bitektine, A. (2011). Toward a theory of social judgments of organizations: The case of legitimacy, reputation, and status. *Academy of Management Review, 36*(1), 151–179.

Bitektine, A., Hill, K., Song, F., & Vandenberghe, C. (2020). Organizational Legitimacy, Reputation, and Status: Insights from Micro-Level Measurement. *Academy of Management Discoveries, 6*(1), 107–136.

Brandts, J., & Figueras, N. (2003). An exploration of reputation formation in experimental games. *Journal of Economic Behavior & Organization, 50*(1), 89–115.

Bromiley, P., Rau, D., & Zhang, Y. (2017). Is R&D risky? *Strategic Management Journal, 38* (4), 876-891.

Carter, S. M., & Ruefli, T. W. (2006). Intra-industry reputation dynamics under a resource-based framework: Assessing the durability factor. *Corporate Reputation Review, 9*(1), 3–25.

Chaiken, S. (1980). Heuristic versus systematic information-processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology, 39*(5), 752–766.

Chaiken, S. (1989). Heuristic and systematic information processing within and beyond the persuasion context. *Unintended Thought*, 212–252.

Chaiken, S. & Ledgerwood, A. 2011. A theory of heuristic and systematic information processing. Handbook of theories of social psychology: Volume one: 246-166.

Chung, K. H., & Pruitt, S. W. (1994). A simple approximation of Tobin's Q. *Financial Management, 23*(3), 70–74.

Connelly, B. L., Certo, S. T., Ireland, R. D., & Reutzel, C. R. (2011). Signaling Theory: A Review and Assessment. *Journal of Management, 37*(1), 39–67.

Cummins, J. D., Wei, R., & Xie, X. 2007. Financial sector integration and information spillovers: Effects of operational risk events on US banks and insurers. Available at SSRN 1071824.

Dierickx, I., & Cool, K. (1989). Asset stock accumulation and sustainability of competitive advantage. *Management Science, 35*(12), 1504–1511.

Fischer, E., & Reuber, R. (2007). The good, the bad, and the unfamiliar: The challenges of reputation formation facing new firms. *Entrepreneurship Theory and Practice, 31*(1), 53–75.

Fried, D., & Givoly, D. (1982). Financial analysts' forecasts of earnings: A better surrogate for market expectations. *Journal of Accounting and Economics, 4*(2), 85–107.

Gatzert, N., Schmit, J. T., & Kolb, A. (2016). Assessing the risks of insuring reputation risk. *Journal of Risk and Insurance, 83*(3), 641–679.

Gilovich, T., Vallone, R., & Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology, 17*(3), 295–314.

Griffin, D., & Tversky, A. (1992). The weighing of evidence and the determinants of confidence. *Cognitive Psychology, 24*(3), 411–435.

Haleblian, J. J., Pfarrer, M. D., & Kiley, J. T. (2017). High-reputation firms and their differential acquisition behaviors. *Strategic Management Journal, 38*(11), 2237–2254.

Hoeppner, B. B., Kelly, J. F., Urbanoski, K. A., & Slaymaker, V. (2011). Comparative utility of a single-item versus multiple-item measure of self-efficacy in predicting relapse among young adults. *Journal of Substance Abuse Treatment, 41*(3), 305–312.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating – the belief-adjustment model. *Cognitive Psychology, 24*(1), 1–55.

Hohle, S. M., & Teigen, K. H. (2015). Forecasting forecasts: The trend effect. *Judgment and Decision Making, 10*(5), 416–428.

Jensen, M., Kim, H., & Kim, B. K. (2012). Meeting expectations: A role-theoretic perspective on reputation. *The Oxford Handbook of Corporate Reputation* (pp. 140–159). Oxford University Press.

Ketchen, D. J., Jr, Ireland, R. D., & Baker, L. T. (2013). The use of archival proxies in strategic management studies: Castles made of sand? *Organizational Research Methods, 16*(1), 32–42.

Lange, D., Lee, P. M., & Dai, Y. (2011). Organizational reputation: A review. *Journal of Management, 37*(1), 153–184.

Love, E. G., & Kraatz, M. (2009). Character, conformity, or the bottom line? How and why downsizing affected corporate reputation. *Academy of Management Journal, 52* (2), 314–335.

MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly, 35*(2), 293–334.

Mitra, D., & Golder, P. N. (2006). How does objective quality affect perceived quality? Short-term effects, long-term effects, and asymmetries. *Marketing Science, 25*(3), 230–247.

Parker, O., Krause, R., & Devers, C. E. (2019). How firm reputation shapes managerial discretion. *Academy of Management Review, 44*(2), 254–278.

Parker, O. N., Krause, R., & Covin, J. G. (2017). Ready, set, slow: How aspiration-relative product quality impacts the rate of new product introduction. *Journal of Management, 43*(7), 2333–2356.

Pfarrer, M. D., Pollock, T. G., & Rindova, V. P. (2010). A tale of two assets: The effects of firm reputation and celebrity on earnings surprises and investors' reactions. *Academy of Management Journal, 53*(5), 1131–1152.

Pollock, T. G., Lashley, K., Rindova, V. P., & Han, J.-H. (2019). Which of these things are not like the others? Comparing the rational, emotional, and moral aspects of reputation, status, celebrity, and stigma. *Academy of Management Annals, 13*(2), 444–478.

Rabin, M., & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *The Quarterly Journal of Economics, 114*(1), 37–82.

Rao, H. (1994). The social construction of reputation: Certification contests, legitimation, and the survival of organizations in the American automobile industry: 1895–1912. *Strategic Management Journal, 15*(S1), 29–44.

Ravasi, D., Rindova, V., Etter, M., & Cornelissen, J. (2018). The Formation of Organizational Reputation. *Academy of Management Annals, 12*(2), 574–599.

Rhee, M., & Haunschild, P. R. (2006). The liability of good reputation: A study of product recalls in the US automobile industry. *Organization Science, 17*(1), 101–117.

Rhee, M., & Valdez, M. E. (2009). Contextual factors surrounding reputation damage with potential implications for reputation repair. *Academy of Management Review, 34* (1), 146–168.

Rindova, V. P., Williamson, I. O., Petkova, A. P., & Sever, J. M. (2005). Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation. *Academy of Management Journal, 48*(6), 1033–1049.

Schmidt, J. A. (2018). Do trends matter? The effects of dynamic performance trends and personality traits on performance appraisals. *Academy of Management Discoveries, 4* (4), 449–471.

Shapiro, C. (1982). Consumer information, product quality, and seller reputation. *The Bell Journal of Economics*, 20–35.

Titus, V., Parker, O., & Bass, A. E. (2018). Ripping off the band-aid: Scrutiny-bundling in the wake of social disapproval. *Academy of Management Journal, 61*(2), 637–660.

Tversky, A. & Kahneman, D. 1974. Judgment under uncertainty: Heuristics and biases. Science, 185(4157): 1124-1131.

Wyer Jr, R. S. 2003. Social comprehension and judgment: The role of situation models, narratives, and implicit theories: Psychology Press.

Wyer, R. S., Adaval, R., & Colcombe, S. J. (2002). Narrative-based representations of social knowledge: Their construction and use in comprehension, memory, and judgment. *Advances in Experimental Social Psychology, 34*(34), 131–197.

Zavyalova, A., Pfarrer, M., Reger, R., & Shapiro, D. 2012. Managing the message: The effects of firm actions and industry spillovers on media coverage following wrongdoing. Academy of Management Journal.

Zavyalova, A., Pfarrer, M. D., Reger, R. K., & Hubbard, T. D. (2016). Reputation as a benefit and a burden? How stakeholders' organizational identification affects the role of reputation following a negative event. *Academy of Management Journal, 59*(1), 253–276.

About the authors

Owen Parker is an assistant professor in the Department of Management at the College of Business at University of Texas at Arlington. His research focuses on how the reputations of firms influence their strategic decisions especially around innovation, as well as gender- and diversity-related nuances in these topics. His work has been published in the Academy of Management Journal, the Academy of Management Review, the Strategic Management Journal, the Journal of Management, Entrepreneurship Theory & Practice, the Frontiers of Entrepreneurship Research, and the Best Paper Proceedings of the Academy of Management. In addition, his work has been presented at the annual meetings of the Academy of Management and the Strategic Management Society, the Oxford University Reputation Symposium, as well as the Babson College Entrepreneurship Research Conference.

Ke Gong is a doctoral student in Management at the Ivy College of Business at Iowa State University. His research sits at the intersection of corporate governance and nonmarket strategy, and his dissertation analyzes how political factors affect the commission and detection of financial fraud. Ke has over ten years of industry experience, including six years as a Senior Business Analyst at Intel China.

Rachel Mui is an assistant professor in the Department of Management at the College of Business at Kansas State University. Her research addresses the intersection of social evaluations, innovation, and gender. Her work has been published in the Strategic Management Journal and presented at the annual meetings of the Academy

of Management and the Strategic Management Society as well as the Oxford University Reputation Symposium.

Varkey (Anand) Titus is an associate professor in the Department of Management at the College of Business at University of Nebraska-Lincoln. His research focuses on corporate venturing, social evaluations, and performance feedback, and has been published in the Journal of Business Research, the Academy of Management Journal, the Strategic Management Journal, the Journal of Management, Entrepreneurship Theory and Practice, and the Frontiers of Entrepreneurship Research. In addition, his work has been presented at the annual meetings of the Academy of Management and the Strategic Management Society, the Oxford University Reputation Symposium, and the Babson College Entrepreneurship Research Conference.

Jiancheng Du is a doctoral student in Management at Old Dominion University. Her research focuses primarily on corporate governance, strategy, and entrepreneurship, and her work has been presented at the annual conferences of the Strategic Management Society and the Academy of Management.

Gyebi Kwarteng is a doctoral student at Oklahoma State University. His research focuses on executive decisions and financial reporting patterns, and his work has been presented at the annual conference of the Strategic Management Society. Gyebi is an accounting instructor at Fort Hays State University and a Business Consultant at Global Citizen Financial. He is a Certified Public Accountant (CPA) and Chartered Global Management Accountant (CGMA) in Illinois. Gyebi has over 15 years industry experience and has worked for several multinational companies such as Baker & McKenzie, Hospira (Now Pfizer), Cardinal Health, AbbVie, and Walgreens in Accounting and Finance roles.