Organization Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Noise as Signal in Learning from Rare Events

David Maslach, Oana Branzei, Claus Rerup, Mark J. Zbaracki

To cite this article:
David Maslach, Oana Branzei, Claus Rerup, Mark J. Zbaracki (2018) Noise as Signal in Learning from Rare Events. Organization Science 29(2):225-246. https://doi.org/10.1287/orsc.2017.1179

http://pubsonline.informs.org/journal/orsc/

ORGANIZATION SCIENCE
Vol. 29, No. 2, March–April 2018, pp. 225–246
ISSN 1047-7039 (print), ISSN 1526-5455 (online)

Noise as Signal in Learning from Rare Events

David Maslach,[a] Oana Branzei,[b] Claus Rerup,[c] Mark J. Zbaracki[b]

[a] Department of Management, College of Business, Florida State University, Tallahassee, Florida 32306; [b] Ivey Business School, University of Western Ontario, London, Ontario N6G 0N1, Canada; [c] Frankfurt School of Finance and Management, D-60322 Frankfurt am Main, Germany

Contact: dmaslach@fsu.edu, http://orcid.org/0000-0002-0631-3695 (DM); obranzei@ivey.uwo.ca (OB); c.rerup@fs.de, http://orcid.org/0000-0002-7222-990X (CR); mzbaracki@ivey.uwo.ca (MJZ)

Received: May 6, 2011
Revised: February 21, 2012; September 14, 2014; October 25, 2016; August 16, 2017
Accepted: September 11, 2017
Published Online in Articles in Advance: April 2, 2018

https://doi.org/10.1287/orsc.2017.1179

Abstract. Firms increasingly have access to information about the failure events of other firms through public repositories. We study one such repository that accumulates reports of adverse events in the medical device industry. We provide qualitative evidence that shows how firms select a sample of adverse events and then engage in inferential learning. We show that firms use the reports of others to extract new valid knowledge from the adverse events in other firms. We use quantitative evidence to explore how a public repository can be used to provide more direct evidence of vicarious learning. Our findings challenge some standard assumptions about vicarious learning. First, we show that the learning in a repository does not come from referent others. Instead, it emerges directly from failure events that might ordinarily be dismissed as noise. Second, we show that the learning does not come from copying others. Instead, it is constructed by firm members as they assemble individual failure events to identify possibilities they had not considered. Third, in contrast to vicarious learning, where the referent others and rare events provide the context, repository-based learning requires that actors impose their own context as part of the learning process. Our qualitative and quantitative evidence serve explanatory purposes by showing how firms use a repository of failure events to identify moments of valid learning, and they serve exploratory purposes by investigating how we can demonstrate reliable learning from a repository of failure events.

Funding: This research project was partially funded by the Canadian Sciences and Humanities Research Council [Standard Grant 410-2010-0325].

Supplemental Material: The e-companion is available at https://doi.org/10.1287/orsc.2017.1179.

Keywords: vicarious learning • learning from failure • rare events • public repositories

Introduction

Firms seeking to learn from failure face a dilemma: as their success increases, the failure experience available for learning diminishes. One solution to this problem is vicarious learning from a public repository of failure events. With the rise of the Internet, firms have greater access to such repositories in a variety of settings, including consumer product safety, computer device failure, disaster and failure events, medical devices, pharmaceutical products, and automobile accidents.[1] Public repositories of failure events expose actors to otherwise unavailable rare events with the hope that it will lead to more effective learning (Lampel et al. 2009, March et al. 1991, Tamuz 2001). Research demonstrates that firms can learn from the failures of others (Baum and Dahlin 2007, Chuang and Baum 2003, Dahlin et al. 2018, Desai et al. 2018, Ingram and Baum 1997, Kim and Miner 2007). Failure data can aid in making more robust inferences by challenging knowledge, recognizing data flaws, or searching for new solutions (Baum and Dahlin 2007, Labib and Read 2013, Miner et al. 1999). Vicarious learning from failure is more likely to occur from other firms that are nearby both geographically and competitively (Kim and Miner 2007). By contrast, a public repository of failure events offers failure reports across diverse firms, products, and locations.

So how can firms learn effectively from a repository of failure events? Effective learning can be assessed by two criteria: reliability, the extent to which an organization develops understandings of experience that are public, stable, and shared; and validity, the extent to which an organization is able to use its learning for understanding, prediction, and control (March et al. 1991). Much of the literature on vicarious learning focuses on reliable learning processes by studying when and how organizations benefit from the knowledge other organizations have acquired (Argote 2013). Firms are more likely to learn from similar and more

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

successful others (Baum and Dahlin 2007, Haunschild and Miner 1997, Labianca et al. 2009). The outcome of these reliable learning processes is a shared, stable understanding, but because actors only imitate success, they are more likely to reproduce existing understandings than to produce new insights (March 2010). Reliability is a necessary but not sufficient condition for learning.

An alternative approach to vicarious learning, learning from rare events (Dye et al. 2014, Kunreuther and Bowman 1997, Lampel et al. 2009), focuses on valid learning processes. Rare events direct attention to salient features of an experience. Rather than sample across a series of similar experiences, actors learn from rare events by focusing on more aspects and interpretations of a single experience. They can then assemble a new and hopefully better causal story from these newly discovered details (Garud et al. 2011, March et al. 1991). Rare events may not produce reliable learning, however, because others may not expect the extreme event to recur and so will choose not to adopt the new valid insights (Lampel et al. 2009, Starbuck 2009). Instead, rare events may be dismissed as noise around the reliable signal from successful existing practice.

A repository of failures lies between these two approaches to learning. The repository takes the knowledge and experience embedded in individuals and organizations (Argote 2013) and makes it publicly available through accounts of their adverse events. Rather than select on the successes of others, the repository intentionally selects on the failures of others. This redresses the tendency to undersample on failure in learning from referent others (Denrell 2003) but by oversampling instead on failure. The adverse reports in a repository also have their limits as a mechanism for vicarious learning. They may be rare relative to high volume production experience (Argote 2013), but the reported events lack the salience of a catastrophic accident (Kunreuther and Bowman 1997, Vaughan 1996). Moreover, the submitted reports are unlikely to match the detailed official accident investigations used in past research (Baum and Dahlin 2007, Haunschild and Sullivan 2002). While the repository may be valuable because it accumulates failures across a very diverse set of firms (Beckman and Haunschild 2002), it can also be seen as the collective noise of an industry.

Here, we study how firms use this noise to find a signal to improve their products and practices. We draw on both qualitative and quantitative evidence on the failure events accumulated in the Manufacturer and User Facility Device Experience (MAUDE) data set of the U.S. Food and Drug Administration (FDA). Any doctor's office, hospital, pharmacy, or veterinarian's office has a vast array of the devices included in this data set: ultrasound equipment, x-ray machines, latex gloves, glucose meters, screws, fasteners, and cutting tools, for example. We use qualitative evidence to outline the inferential learning process by which firms establish new valid inferences from the details of individual failure events in the MAUDE data set. In our quantitative evidence, we explore how a repository of failure events can be used to demonstrate vicarious learning. We find that vicarious learning from a public repository of rare events challenges some standard assumptions about vicarious learning. First, the signal from a repository cannot be drawn from the successful practices of others; instead, the signal emerges from the failure events that ordinarily would be dismissed as noise. Second, that signal also is not drawn from aggregate experience; rather, it is constructed by firm members as they assemble specific failure events to identify possibilities they had not considered. Third, in contrast to vicarious learning, where the referent others and rare events provide the context, repository-based learning requires that actors impose their own context as part of the learning process. Our qualitative and quantitative evidence serve both explanatory and exploratory purposes (Creswell 2003): explanatory purposes by showing how firms use a repository of failure events to identify moments of valid learning, and exploratory purposes by investigating how we can use quantitative evidence to demonstrate reliable learning from a repository of failure events.

Signal Extraction in Organizational Learning

The trade-off presented by a public repository of failure events is an instance of a general problem: actors cannot attend to all available experience (Ocasio 1997, 2011). Learning from referent others, learning from rare events, and learning from a repository all direct attention in different ways and involve different trade-offs for learning. In learning from referent others, firms direct their attention to successful and similar others, but they oversample on success (Denrell 2003). In learning from rare events, firms direct their attention to the details of a single event, but extreme events are challenging to interpret and learn from (Lampel et al. 2009). A repository of rare events presents an abundance of failure events, but how do firms select from the repository a sample of events from which they can expect to learn? The three forms of learning involve choices about what experiences are most likely to lead to effective learning.

Vicarious Learning from a Reference Group

The dominant approach to vicarious learning assumes that firms compare their performance to competitors and engage in vicarious learning when they fall short of performance targets (Greve 2003). Evidence shows that vicarious learning will be easiest and most likely to occur with similar and more successful referent

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

others (Kacperczyk et al. 2014, Lieberman and Asaba 2006, Massini et al. 2005, Moliterno et al. 2014, Strang and Macy 2001). Firms determine that reference group when they structure their competitive space. For example, strategic groups form a cognitive community that informs managerial action (Porac et al. 1989). When seeking new knowledge, firms typically choose the higher-performing firms as targets for imitation (Baum and Dahlin 2007, Labianca et al. 2009).

Learning from referent others may not always come from direct evidence about what causes success and failure. Instead, firms—and researchers—use performance data that are "visible and salient, interpretable . . . and generalizable across organizations" (Baum and Dahlin 2007, p. 370) to demonstrate a reduction in failure rates. As Madsen and Desai (2010, p. 453) observe, such measures necessarily offer only indirect evidence of learning: "Given the difficulty of observing changes in organizational knowledge itself, the assumption in much of the empirical organizational learning literature is that changes in observable organizational performance reflect changes in organizational knowledge."

Learning from referent others occurs when firms engage in "trait imitation" (Haunschild and Miner 1997, p. 472) by copying their most successful peers. The result will be a generally conservative approach to learning, reproducing a consistent set of generally effective practices, but ignoring contradictory information. As a result, the behavior in a competitive set (Porac et al. 1989) may become more similar without necessarily becoming more valid. This is one consequence of oversampling on success (Denrell 2003). Trait imitation can be used to adopt effective practices, but it can also lead firms to adopt ineffective practices (Abrahamson 1991, DiMaggio and Powell 1983).

Vicarious Learning from Rare Events

Rare events are a possible antidote to this biased sample. Reference-group learning focuses attention on the successes of others, but rare events focus attention on more aspects of a single experience (Lampel et al. 2009, March et al. 1991). Rare events such as a space shuttle explosion (Vaughan 1996) or a chemical plant explosion (Kunreuther and Bowman 1997) can serve as branching points in history. For example, the Union Carbide Bhopal accident changed the risk preferences and reference points (March and Shapira 1992) for other firms and gave them reason to improve oversight, reduce toxic inventories, develop evacuation plans, and change standard operating procedures (Kunreuther and Bowman 1997).

Learning from a rare event requires determining whether an event has lessons for preventing future occurrences (Lampel et al. 2009). With rare events, this is a difficult task. Individuals cannot attend to all features of a given event (Hoffman and Ocasio 2001), so an unlikely event may be dismissed as an outlier. If actors do not dismiss the event, they must then select aspects of the event to carry into their actions. However, actors may engage in superstitious learning (Denrell 2008, Zollo 2009) and use observed outcomes to justify changes that are driven by spurious, rather than causal, relationships (Lampel et al. 2009). In addition, even valid lessons might not improve behavior. A rare event may produce transitory lessons that are written off when the world returns to its previous state (Garud et al. 2011, Lampel et al. 2009). Valid learning may lead to overconfidence and more risk taking (March and Shapira 1992, Zollo 2009), so even valid inferences from a rare event may not produce reliable learning processes.

Vicarious Learning from a Repository

A public repository of failure events redresses the biases toward the success of a reference group and the limitations of a single rare event, but the repository also brings its own set of challenges. First, how do firms choose a sample of experience from a repository? A public repository such as the MAUDE data set provides failure reports that cut across strategic groups and cognitive communities (Porac et al. 1989, 1995), so choosing a sample based on reference groups will not work. Because the repository provides only direct reports of failure events rather than estimates of production or usage volume, firms cannot calculate failure rates to establish performance measures. Firms could use proxy measures to construct a reference group of comparable and more successful others (Greve 2003, Kacperczyk et al. 2014), but that would substantially limit the range of data a firm might use.

Second, how do firms extract a signal from that sample? The repository contains only failure reports, so the accumulated data now oversample on failure rather than success. Furthermore, the variety of sources may produce reports of varying quality. The repository offers an increased volume of reports but brings with it the collective noise of an entire industry, replete with vividness problems, media biases, and a range of other forms of bias that may inhibit learning (e.g., Denrell and Le Mens 2016). Standard models of learning would dismiss this as noise rather than treat it as the basis for learning.

Third, how do firms engage in effective learning from a repository of failure events? Existing models of learning draw statistical inferences from cumulative and comparable experience (Bohn 1995), but to learn from a repository of rare events, a firm must find other ways to process and combine information:

> Moving from rich experiences of history to valid inferences about history involves a logic that is not well-defined but is different from the logic of classical statistical inference. It assumes that the various micro events

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

> associated with an event are some way interconnected. They are clearly not independent samples of some universe in the standard statistical sense. But they provide scraps of information about an underlying reality that cumulate, much the way various elements of a portrait cumulate to provide information about its subject.
>
> (March et al. 1991, p. 8)

This is not typical vicarious learning, but a repository such as the MAUDE data set intentionally shifts attention away from reference groups and single rare events in the hopes that firms can learn from the failure events of others. Learning from a reference group privileges the traits of similar and successful others but at the cost of reinforcing ideas that are not valid. Learning from rare events has the potential to develop new valid inferences, but single events may not lead to reliable learning processes (Lampel et al. 2009). In comparison, a repository offers reports across a range of events. Firms can use the failure reports to challenge existing causal understandings. If learning is to be effective, however, the firms must create new valid understandings that can then become public, stable, and shared. In subsequent sections, we explore the process by which firms engage in such learning.

Qualitative Evidence on Learning from the MAUDE Data Set

Qualitative Methods

A study of an emerging approach to learning must present valid evidence that reflects the actual learning process (Kim and Miner 2007). This problem is particularly acute in vicarious learning from a repository such as the MAUDE data set, which expects that users can extract a signal from an immense set of rare events drawn from a very diverse set of firms. As our theoretical overview indicates, this requires a different form of learning. Following recent mixed-methods approaches (Turner et al. 2017), we gathered qualitative data to support our hypothesis development and interpretation of our results. Following standard mixed-methods research approaches (Gibson 2017), we gathered our qualitative data after our quantitative data, but because repository-based learning is a theoretically new area, we begin by presenting the qualitative evidence to show how individuals select adverse events and draw lessons from those events. (Here, we give an overview of our methods; more details can be found in Online Appendix A.)

We used purposeful sampling to gather our qualitative data. We first sent a survey to all device manufacturers registered with the FDA. These are the firms that would both contribute to and learn from the MAUDE data. We received 190 responses; 107 of those firms indicated interest in a follow-up interview. From that list, we purposefully selected firms with a range of failure experience and with different sizes and structures, including small family-owned firms and large, public multinationals. We also sought firms from a variety of product markets, including medical imaging; ear, nose, and throat; neurology; and other device markets. From our sample of firms, we chose 31 informants to represent a diverse range of adverse event experience, occupation, and roles, including chief executive officer or president, middle management, regulatory analyst, sales, device design, and consultant roles. Of those 31, 12 agreed to participate. Our semistructured interviews varied in length from 45 to 120 minutes. We also conducted multiple follow-up interviews with our informants. All interviews were recorded and transcribed verbatim. Our interview protocol used a range of open-ended and directed questions that would avoid demand effects while still eliciting evidence of learning from adverse event reports in general and the MAUDE data set specifically. We also used site visits to manufacturers, hospitals, and research labs; long-standing relationships with a range of medical device executives; and a 150-person conference on medical device innovation organized by the first author to inform our qualitative methods. Finally, we drew on media accounts of medical device adverse events drawn from the MAUDE data set to identify how adverse events led to reliable and valid learning processes.

Inferential Learning from a Public Repository

The website for the MAUDE data set includes an important disclaimer:

> Although MDRs [medical device reports] are a valuable source of information, this passive surveillance system has limitations, including the potential submission of incomplete, inaccurate, untimely, unverified, or biased data. In addition, the incidence or prevalence of an event cannot be determined from this reporting system alone due to potential under-reporting of events and lack of information about frequency of device use. . . . MDR data alone cannot be used to establish rates of events, evaluate a change in event rates over time or compare event rates between devices. The number of reports cannot be interpreted or used in isolation to reach conclusions about the existence, severity, or frequency of problems associated with devices. Confirming whether a device actually caused a specific event can be difficult based solely on information provided in a given report. Establishing a cause-and-effect relationship is especially difficult if circumstances surrounding the event have not been verified or if the device in question has not been directly evaluated. (U.S. Food and Drug Administration 2016)

This disclaimer highlights the challenges of building a reliable learning process from failure events in a public repository. The MAUDE data suffer from some significant limitations that stand in the way of constructing and sharing stable beliefs about medical device failures. The quality and consistency of the reports is

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

suspect, so it is hard to reconcile the reports with existing stable understandings. The reports identify diverse events that are often unique, so clear comparisons are not always evident. There is no information about failure rates, so it is also impossible to identify clear referent others and learn by copying their practices. And the MAUDE data set focuses only on failure events, but firms with more failure experience are often less capable and hence poor targets for learning.

The challenge with this noisy data is identifying the failure events from which the firm can learn. One informant, a retired regulatory consultant for numerous medical device firms, explained the challenges:

> Man, I went through the MAUDE data set looking for all neurological devices, all needles, and all electrodes, and I found nothing that was similar. This was a real unicorn. I really did comb quite thoroughly. Plus, I looked at other needles in other categories. You look at "who" they were reported against. You have manufacturers in there and you see if they have any reports against them. I am a pretty good "Googler [searcher]." Those are the skills you need to sort through a mass of data like this. To ferret out what's in there. You look at anything that came close to . . . [the problem].

As the informant points out, using standard learning processes with the MAUDE data will not work. The informant used the MAUDE data to identify and compare similar devices—neurological devices, needles, and electrodes—and found nothing. Yet the informants said that they found the data useful for actively identifying a range of relevant industry details. As a senior manager of U.S. regulations for a large medical device multinational pointed out, the MAUDE data set is part of a broader trend around publishing and using a broad range of information in the medical device industry:

> I do a lot of digging on websites and clinicaltrials.gov to see if anyone is doing clinical studies. . . . People with websites are giving more and more information now. Before it was really hard to find people's user guides and things like that, but a lot of people in the aesthetic world have started posting stuff linked from their website on a YouTube site or Instagram. Social media really has allowed this different level of sharing of knowledge. Whether they are reposting something a doctor did or sharing patient's success stories on Instagram or Facebook. It's a really great window into seeing the market that they are targeting, the claims they are targeting, from an ad/promo value—things like that. The need to have a lot more interaction on websites has driven device companies to put more out there than say five or ten years ago when you tried to make sure your competitor couldn't get a lot of stuff to use for their own purposes.

How do firms make sense of these details? First, they use thresholds of occurrences to determine whether a particular type of event needs to be incorporated in their inferential learning processes. We found multiple instances when informants referred to thresholds, baselines, norms, and benchmarks in choosing to attend to adverse events—either in their own experiences or those of other firms. These thresholds did not necessarily mean that they sought large volumes of events. Instead, our informants sought rich detail about more aspects of others' experiences. Rather than aggregate the number of events, the informants took these multiple pieces of information and assembled them into a pattern. For instance, an increase in the volume of adverse event reports was insufficient by itself. The informants needed to identify a pattern in the increase. As a regulatory specialist at a focused healthcare multinational said, a spike was only a trigger for potential learning: "If we do see a spike we . . . go back and say what is going on there, what do we need to understand this problem?" Learning from a sample of adverse events required identifying the new and unusual details of the events that could inform new valid inferences.

This approach is consistent with a learning process that focuses more on making valid inferences from data. Instead of pooling across events, an inferential learning process involves accumulating and pooling features of a sample of events and assembling a pattern from the available scraps of knowledge. From the pattern, our informants could develop and use a causal model to anticipate failures and make changes to their own devices. A critical task was therefore selecting events from which valid learning was most likely to occur.

Event Selection

Not all learning from adverse events required the MAUDE data. Some events were obvious because they galvanized industry attention. For example, as an executive at a small private medical device firm said, "If one of my competitor's products killed one of their [patients], boy, we'd just all know about it overnight." These stories traveled swiftly, so the MAUDE data set was important only to allow actors to garner the specific details of the incident. In addition, while search on competitors was possible, our informants did not see competitor events as learning opportunities because they likely already knew about those events. Instead, they more often used the MAUDE data set to identify events that would reveal new possibilities. This form of inferential learning required more careful thought. Rather than draw the sample from either a reference set or a significant event, our informants followed two sampling strategies in working with medical device failures, selecting on either failure modes or protocol violations.

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

Failure Mode. One common sampling strategy was to search the MAUDE data set for failure modes: device use that led to a specific category of adverse event. Most device manufacturers already understood the ways that a device might fail and had identified many of the possible causes. The MAUDE data set allowed them to explore unexpected causes and outcomes of a failure. One approach was to compare the firm's own failure experience against that of others to see what was causing the failure. For example, one senior manager that we interviewed told us, "Many times, you target an event, like burns or nerve damage, where you are trying to target a topic and not be specific to a competitor, to see if it is more of an industry thing." As the manager explained, a focus on a specific failure mode could be used to compare a tentative lesson from the firm's own experience against the experience of others to see whether the problem was consistent across actors in the industry. For instance, a large multinational company identified a possible problem of misuse with its devices and used the MAUDE data to see whether multiple other firms experienced similar problems of burns or nerve damage events from misuse. A senior manager from the firm explained, "We went online to specifically look at that outcome and it was very clear that other competitors were having very similar issues of misuse of the device." In this instance, managers used a specific failure mode frequency to see whether the failures should be attributed to the firm's device alone or whether the failures were common to the device class. The informants began with a new interpretation—misuse of its device—and then referred to the MAUDE data to see whether the interpretation could be used to predict outcomes across actors.

Another use of failure modes as a sampling strategy involved selecting on critical events to explore unexpected causes. A senior manager of preclinical affairs gave an example: "Those devices, say, like blood contacting devices (like a cardiovascular stent), when things go wrong, they kind of go wrong *epically*." In these instances, the actors began with a potential failure mode and used the adverse events to identify causes that they had not expected. The senior manager explained,

> Say, for example, there's something in a technology that if it fractured or something it could penetrate a critical organ and cause death. These types of things you want to make sure you understand the potential incidents of that type of scenario, and any type of event that might occur theoretically. Then do whatever amount of testing is required to basically design that out of the product as possible.

This sampling strategy is a form of enacted salience in which the significance of the event drives the inferential learning. The events could have severe consequences, so the informants used the MAUDE data to identify possible causes of failure that they had not considered. The informants found events that had not happened to them, but could happen, and used those events as the basis for vicarious learning.

Sometimes a single unique event could lead to inferences that overturned existing shared understandings. MAUDE Data Report 2183926-2016-00479 describes one example, a single instance when a computer monitor went blank during a heart catheterization procedure. The report revealed that a personal computer had lost contact with the device that had been inserted in the patient to monitor physiological data during the procedure. The team performing the procedure had to wait five minutes with a sedated patient while the computer rebooted the application. The firm's investigation found that the antivirus software performed its hourly scan during the procedure, causing the catheterization monitoring application to crash. The first media report disseminating the incident extended the lessons from this event to any software-enabled equipment: "Just like any other software package, Merge Hemo is subject to the same limitations and dangers that other applications face, and sometimes may crash" (Cimpanu 2016). Another media report demonstrated how the new valid inferences could lead to vicarious learning across a range of actors: "This Internet of Medical Things doesn't just include the aforementioned cardiac catheter machine, it can also include implantable devices like a pacemaker or a brain implant—devices that will more surely kill you if they needed five minutes to reboot" (Ehrenkranz 2016). A manufacturer of an implantable device would not compete with a manufacturer of devices to monitor patients during a medical procedure, but the failure modes could be common to each. As one informant commented, "Adverse events aren't unique to each company or device class; it is just going to be everywhere."

Protocol Violations. The second sampling strategy was to use MAUDE data to select protocol violations: the failure to follow proper procedures. Protocol violations could be used to draw attention to either accidental or intentional noncompliance with approved or expected procedures. For instance, one informant told us how he looked for what he called "precedents," which included decisions made by other firms, users, or the FDA. These precedents could then be used to explore noncompliance as a precursor to failure. For example, an informant used FDA warning letters to sample experience:

> You'd find lists of people who had warning letters . . . typically a huge number of warning letters are to do with the way people market things. You can get an idea of what people are doing, what the FDA doesn't like, what you can do in the future.

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

These letters helped informants identify instances in which the FDA was particularly concerned about the possible increased risks or diminished benefits from a behavior.

Protocol violations could also be used to identify unintended or unexpected device uses. For instance, one informant described how she looked for "off-label" use of her firm's devices—instances in which someone used the device in ways that the manufacturer did not intend. Such protocol violations could be used in combination with a specific failure mode to see if an adverse event was a consequence of off-label use. The senior manager looking for nerve damage also used protocol violations to identify adverse events. He said that the concern emerged from a discovery of off-label use: "We've had people more recently use the device off-label; with that, it was leading to a specific outcome." In this instance, the firm learned "that it obviously wasn't just our device that the people are misusing."

Protocol violations could also be used to flag unusual reports. One common form was a delayed report. The MAUDE data set requires a report be filed within 30 days, so a delayed report was unusual. A delay could simply be an outlier, but it could also signal a possible branching point from which vicarious learning was possible. A director of regulatory affairs for a large multinational enterprise (MNE) explained,

> The delay is one of two things. . . . [First,] it is a human process error because we have 30 days to report an adverse event; it shouldn't happen. Even if you find out it's late, you still go ahead and report. And then try to put a CAP [corrective action plan] in place so that doesn't happen again. [Second], when the event occurred, there really was no injury; no one was hurt. And you didn't have any information to believe that somebody would be hurt if it occurred, but maybe you still continue investigations along the way. After your investigation, you . . . have more information, and . . . believe that we should communicate this information. The awareness date is the date that you became aware of the new information that suggests you may have the potential for an adverse event. And not necessarily the original report. The original report probably would trigger some kind of investigation. And maybe as part of that initial investigation you found other things that were not apparent in the first report. And it became only apparent to you after your investigation. And in that case, your event date is the date that technically you became aware of the potential MDR. So it may appear too late, but it's really not "late."

As this explanation shows, understanding the protocol is essential to determining whether a late report was an outlier—for example, reporting errors—or a sign that the device manufacturer had uncovered new and important information that could be used for learning.

If the delay conveys new information, then the noise in the data set becomes a feature rather than a bug in the learning process. Reliable learning processes protect established beliefs by focusing on the standardized history, but by ignoring anomalies and missing data, they also suppress new and unique inferences that could extend existing knowledge. By contrast, the nonstandard reports revealed uses of medical devices that the informants did not expect. For example, a senior manager of a large healthcare multinational enterprise described how focusing on off-label use helped to identify patients as another potential source of knowledge:

> If you look at a chain [of beauty clinics] that is in multiple states and they were super high volume. . . . It was definitely quantity over quality. . . . I think that is the thing with these devices that our consumer base—that your doctor is not necessarily the one reporting the adverse event. Many times it is the patient calling and saying, "I have bruising, burns, pitting in my skin," because they are unhappy many weeks down the road. Because they leave, and they think they are going to look like 20 again. But they don't. Some of these events are about misuse. I think that also drives a different impact of reporting in some of these cases. . . . Patients tend to report things more than doctors. For invasive surgery, like liposuction, it's hard to get doctors to report it appropriately.

Although doctors are unlikely to reveal careless use or misuse of a device, the patients will. Even apparently unreliable reports that might be dismissed as noise can signal learning.

Inferential Learning from a Repository: Analytic Summary

Existing research on organizational learning uses changes in observable outcomes as a proxy for measuring changes in organizational knowledge (Madsen and Desai 2010). Typically, this work assumes that firms learn by making comparisons to similar or successful others (Baum and Dahlin 2007). By contrast, the MAUDE data offer the possibility of demonstrating more direct evidence of vicarious learning. Rather than looking across a large sample of experience from similar referent others and then copying behaviors that are presumed to lead to success, our informants selected adverse events from which they thought learning would be most likely to occur. They selected either on specific failure modes or on protocol violations. The adverse events they chose could be used to identify aspects of experience that they had not seen in their own experience, to find more ways of seeing those adverse events, and to learn from events that would not have happened with their products. If they made comparisons to referent others, the referent firms were chosen based on lessons already drawn from the adverse events. They would then compare across a (more limited) sample of events so that they could establish that

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

their predictions were reliable across actors, products, or time.

In our qualitative evidence, we have substantiated that firm members use the failure events of others to develop valid learning processes. The next step in an inquiry into effective learning from a repository would be to determine whether repositories lead to reliable learning processes. To do this, we need to use quantitative evidence to aggregate across firms and adverse events to show that inferential learning can lead to changes in observable outcomes.

Quantitative Evidence on Learning from the MAUDE Data Set

> How come two big companies that make the exact same product may not have one common adverse event? It seems as though what some company considers as an adverse event, the other company doesn't. And it's not unusual. . . . [Does] that mean they don't have any? That's impossible. They probably do, but they don't believe [it] . . . . But how can we be really more objective?
>
> (Director of regulatory affairs for a large MNE)

The adverse events in the MAUDE data set are all potentially critical events (March et al. 1991) that may replace existing knowledge with a new understanding that leads to better prediction and control of outcomes. The modeling challenge is how to identify which events are most likely to lead to learning. As the quote illustrates, our informants engaged in vicarious learning by comparing different definitions of an adverse event. If a definition reflects a public, stable, and shared understanding, then it should be replicated in the adverse event reports of others. Following our qualitative evidence, we extend that logic by selecting one specific protocol violation, reporting delays, as the basis for identifying events from which firms might learn. With a more homogeneous data set, it would make sense to select on all failure modes or all protocol violations, but because devices are so diverse, it is impossible to identify the range of possible failure modes and protocol violations. To focus on learning from adverse events, we construct measures to demonstrate how failure events—rather than failure rates—produce evidence of learning. We use those measures to show how reporting volume and delays affect learning from the MAUDE data.

Hypotheses

Modeling and testing the MAUDE data learning processes must begin without identifying firms as referent others. A public repository of failure events encourages learning across a broad range of firms and medical devices. Beginning with a set of comparable others would unnecessarily restrict the range of contexts available for learning, so we follow the learning process and induce the reference set from the evidence of learning from adverse event reports. In inferential learning, the reference group forms after firms construct understandings of an adverse event, the causes of that adverse event, and solutions to overcome that adverse event. If vicarious learning is happening, the new understandings will be adopted by firms that use similar technologies, face similar technical challenges, or make similar use of a device. Learning should lead them to report similar failures. This process focuses on failure *reports*, not failure *rates*. Standard approaches to vicarious learning use failure rates—the number of failures over experience—to demonstrate learning as failure rates decline (Baum and Dahlin 2007, Kim and Miner 2007). In our setting these failure rates may vary dramatically because the frequency of use may vary dramatically relative to the number of devices (e.g., a computed tomography (CT) scanner versus a deposable syringe), so experience-based failure rates may be a poor proxy for lessons extracted. Instead, we should expect a pattern of failure reports to follow initial reports of a specific adverse event. If such inferential learning happens, we can derive the reference set by looking for firms with similar patterns of failure reports.

The vicarious learning process should proceed in two steps. First, potentially valid lessons emerge through the reports of a focal firm. Second, the set of relevant others—those who find some merit in the adverse event report—will then report similar failure events in their own device use. As others adopt those lessons, the understanding becomes public, stable, and shared. We should then see an increase in the number of reported adverse events as the common definition takes hold and the other firms submit reports. The second step is essential to demonstrating that this is an accurate understanding of reality. If the new shared definition offers knowledge that can be used for prediction and control, then we should expect to see a decline in adverse event reports in the subsequent period. Hence, we predict the following.

Hypothesis 1 (H1). *If a focal firm has demonstrated a pattern of adverse event reports that correlates with the pattern of adverse event reports of a set of other firms, then an increase in the number of adverse event reports by that set of relevant others in time period t will lead to a decrease in the adverse event reports by that focal firm in subsequent periods $t + n$.*

Volume as Inhibitor. The first hypothesis establishes a baseline of learning from relevant others as determined by failure events but does not show how firms select the sample of events from which to learn. Firms need to select a sample of experience from the data set that offers a valid inference. In a learning process that relies on referent others, the selection process is simple: choose those who do similar work and produce

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

fewer failures (Kacperczyk et al. 2014, Lieberman and Asaba 2006, Massini et al. 2005, Moliterno et al. 2014, Strang and Macy 2001). The selection criterion works because the learning process in the referent others follows reinforcement learning: success leads to continued behavior, while failure leads to abandonment (Cyert and March 1963). Vicarious learning follows as firms copy the insights gained by referent others with similar experience. That experience can be measured in a variety of forms—ship tonnage produced (Argote et al. 1990), flight departures (Haunschild and Sullivan 2002), operating miles (Baum and Dahlin 2007), and orbital vehicles launched (Madsen and Desai 2010).

A repository of adverse events is different because it includes only failure reports. There is no denominator to measure the rate, and there is no information about the context of the failure, so volume alone is an ambiguous signal. An increase in the volume of failure events could simply follow from an increase in device production volume. An increase could also signal that the firm is reporting more failures because it had discovered new information that would offer opportunities for vicarious learning. Or, an increase in the volume of adverse events could signal normal variation in failures—spurious relationships rather than causal relationships (Greve 2003, Lampel et al. 2009). In this last instance, other firms may respond in one of two ways. First, they may engage in superstitious learning (Denrell 2008, Levitt and March 1988, Zollo 2009) and respond to the spurious relationships. Second, they may take an increase in failures as a sign that the firm does not know what it is doing and ignore a valid signal of a causal relationship (Desai 2015, Diwas et al. 2013). In combination, these two errors in learning mean that an increased volume of adverse events will create noise in the system and inhibit learning. This prediction is for changes based on volume increase alone—that is, situations in which firms have responded to adverse event reports without engaging in the careful inferential learning described in our qualitative findings. Hence, we predict the following.

Hypothesis 2 (H2). *As the volume of adverse events from other medical device manufacturers increases, the learning from these adverse events by the focal firm will decrease.*

Delays as Inhibitor. Delays are an ambiguous signal (Argote 2013, p. 37). As observed previously, delays are a protocol violation and may be grounds for sample selection. The FDA mandates that firms report adverse events within 30 days of learning of an incident (Medical Device Reporting, Manufacturer Reporting Requirements, 21 C.F.R. 8 (2015)). The FDA can—and does—impose penalties on firms that fail to file timely reports (Delporte 2011), so an adverse event report should arrive close to the time of an incident. Reports that violate this mandate may signal unusual events that merit attention.

Like volume increases, reporting delays also introduce noise that will inhibit learning. This prediction is consistent with the existing literature that shows that reporting delays are likely to make learning more difficult (Denrell et al. 2004, Rahmandad 2008). There are two fundamental reasons that this should hold for the MAUDE data. First, firms may conclude that delayed reports cannot be trusted. Given the reporting requirement, other firms would expect a firm experiencing an adverse event to submit a report even if the cause of the incident is not yet clear. A delayed report may signal human processing error, a lack of understanding on the part of the reporting firm, or that the reporting firm is hiding something. As a result, firms may dismiss legitimate causal evidence from those reporting firms.

Second, a failure experience is not available to other firms until the report is filed. Much of the learning from adverse event reports comes from the assembly of scraps of information (March et al. 1991). Without direct attention close to the time of an adverse event, contextual information behind an event is easily forgotten (de Holan and Phillips 2004, Thompson 2007), and firms may miss, forget, or ignore critical information. As delays increase in duration, these effects should increase, as the attention of the firms that did not experience the adverse event will go to other sources of learning, including either their own experience or other reported adverse events (Ocasio 1997, 2011). These arguments are supported by empirical research that shows how delays between cause and effect substantially limit learning (Diehl and Sterman 1995, Repenning and Sterman 2002). We therefore predict the following.

Hypothesis 3 (H3). *As the delays in adverse event reports filed by other medical device manufactures increase, the focal firm's learning from these adverse event reports will decrease.*

Delays and Volume as Signal. Vicarious learning from a repository of failure events depends on identifying adverse event reports that can be used to establish valid inferences. In the previous two hypotheses, we argued that increases in the volume of failure reports and reporting delays introduce noise that inhibits vicarious learning. From our qualitative evidence, we argue that in certain situations, delays and volume can interact to make possible valid inferences from the reports of others. As noted above, reports filed after the FDA-mandated 30-day reporting deadline are protocol violations. Subsequent reports can occur more than a year—or multiple years—later. A delayed report may inhibit learning (Rahmandad 2008), but a delay may also signal valid lessons (Li et al. 2013, Strike and Rerup 2016). Long delays may signal intentional protocol

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

violations for infrequent adverse events or for adverse events in which the consequences appear long after the medical device was used. For example, incidents such as corroding leads on heart pacemakers have serious consequences, but only after long-term use (Geddes 1995). Other delays may occur when investigations reveal adverse events that initially went unnoticed or must be reclassified. Delays may also occur when firms discover unexpected interactions with other medical devices. Here, it is important to distinguish the timing of the arrival of a report—which inhibits learning because the report is not available—from the signal that a delayed report sends. Since the adverse event reports are part of a repository, firms may not process them in real time. Instead, they may query the database and draw up a set of reports that were filed in the past. In these instances, the delay signals the rich learning through which the firm has identified omitted variables or alternative interpretations (March et al. 1991) of an adverse event. If a spike in volume of delayed reports points to a potentially important failure mode, then we should expect firms to give these delayed reports particular attention. As a result, for high volumes of adverse events we should see the inhibiting effects of delays attenuated. Therefore, we predict the following.

Hypothesis 4 (H4). *The inhibiting effect of delays on learning will be larger for a small volume of adverse events than for a large volume of adverse events. Thus, for large volumes of adverse events from other medical device manufacturers, the inhibiting effect of delays on learning by the focal firm will diminish.*

Quantitative Methods

Data Collection. To test our hypotheses, we built a data set of all medical device adverse events in the MAUDE data set in a 15-year period, from January 1, 1997, to November 30, 2012.[2] There was an upward trend in adverse event reporting as a result of regulatory requirements introduced in 1997. To match adverse event reports with medical devices, we merged mandatory reports of adverse events recorded in MAUDE with the FDA's firm-level registration database and the medical device registrations in the FDA's 510(k) Premarket Notification database. The 510(k) database delineates each firms' intent to sell a medical device in the United States. The population of manufacturers is drawn from the Establishment Registration and Device Listing data set. The registration data set includes no data on inactive firms, but the MAUDE data set does contain inactive firms and their products. To minimize any censoring problems that this may cause, we focus on the 5,343 parent firms from the registration data set in 2010. Of these, 586 parent firms (21,592 firm-month observations, 11% of the population) reported at least one adverse event, and together they reported a total of 1,064,364 adverse events from 1997 to 2012. These firms are manufacturers, contract manufacturers, distributors, and firms that repack or relabel existing devices.

The merged data set gave us failure reports from a broad range of products and firms in the medical device industry. Most studies of learning use clearly defined referent others producing similar products (e.g., Argote et al. 1990). As our qualitative findings show, firms using the MAUDE data can learn from the rare events of firms producing very different products. For purposes of demonstrating quantitative evidence of learning, this variability presents analytic challenges. The reported events are not rare enough to be samples of one or fewer, but they are sufficiently rare that many firms do not report an event in a given year. Adverse events from medical devices can have dire consequences, so the rarity is desirable, but it is hard to show evidence of learning for firms that do not report adverse events. Dealing with the firms that do not report is important to our sampling strategy because those firms can drive the results in ways that undermine our analysis. There are three broad approaches to this challenge.

One approach is to include the entire population. This least restrictive approach reflects the inclusive spirit of the MAUDE data set, but it presents significant problems because the failure data are driven by the long tail of rare events (Starbuck 2009). Many medical device firms report only once or not at all, and when they do, they often report clusters of similar problems. The vast majority of these rare events truly are simply noise: they reflect idiosyncratic experiences that other firms cannot learn from. In addition, because many of the firms do not experience failures in every period, the large sample of mostly zeros means that the actual outlying events receive extra weight in the analyses, causing collinearity problems. Demonstrating vicarious learning between firms requires dynamic comparisons from each individual firm to all other firms. For firms with no failure events, it is hard to find evidence of learning.

A second alternative is to create a more restrictive population focused on firms that clearly demonstrate learning. For example, establishing a minimum threshold of reports over the course of the sampling window would select on firms that are clearly regularly reporting adverse events. The evidence from these firms could then be used to demonstrate that a focal firm monitoring reports in the MAUDE database will reveal the effects in its own subsequent reports. These sampling criteria are not entirely consistent with the inclusive spirit of the MAUDE data set, though they still include a broad range of firms. However, without a theoretically grounded minimum threshold for failure events, the cutoff for this sample would appear to be arbitrary.

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

A third possibility, and the one we follow here, is to focus on the firms that most consistently report adverse events: the sample of 13 firms that reported adverse events in every month for 15 years from 1997 to 2012 (Alcon, Abbott, Boston Scientific, Cordis, C. R. Bard, Depuy, Ethicon, Invacare, Medtronic, Smith & Nephew, St. Jude, Teleflex, and Zimmer). Standard approaches to vicarious learning would choose the firms with the most experience, under the assumption that those firms are most likely to be learning. Because our data do not provide evidence on actual device use, this experience is hard to measure. Some firms may sell only a few devices that are used in a large number of procedures (e.g., CT scanners). Reporting frequency is the most appropriate alternative and should yield a similar set of firms because more experience is likely to yield more adverse events. The 13 firms work on a diverse set of technologies and are industry leaders. This more tractable sample lends itself to a richer exploratory analysis while still drawing on a sizable database of failure experience that accounts for 44% (466,894) of the adverse events in our population. The final sample was a balanced 15-year panel of 2,483 firm-month observations (191 months), which spanned 19 product sectors and 1,253 product classifications.[3] Note that our findings are robust across all three alternative sampling strategies. (See Online Appendix B for descriptions, robustness checks, and the trade-offs in analysis for the more inclusive samples.)

Dependent Variable

Our primary dependent variable, $R_{i,t}$, is the number of adverse events reported by a firm i from the current month (t). Unlike prior operationalizations of accident and incident rates (Haunschild and Sullivan 2002), our dependent measure is event based, so it is not scaled.

Independent Variables

Vicarious Learning. Because firms are heterogeneous and firm learning rates differ (Lapré and Tsikriktsis 2006, Pisano et al. 2001), our null hypothesis is that there exist no similar failure patterns across firms. Rejecting this null hypothesis implies that systematic patterns of failure reductions exist across the failure trajectories of firms, which are the firm-specific paths at which failure experiences change over time, providing evidence for vicarious learning between firms.

The measure of reliable vicarious learning between firms in period $t-1$ ($E_{j,t-1}$) is the average interfirm correlation between a firm's failure trajectory and another firm's failure trajectory in the past. It is constructed in two stages. The first stage calculates the slope of the failure trajectory for a firm—the change in failures per month due to past failure experience over time. The second stage estimates the correlation between the firm's trajectory and all others. A significant coefficient suggests that failure patterns exist between firms and that these patterns reflect either a reduction in adverse events (a negative coefficient) or an increase in adverse event reporting (a positive coefficient). Note that either direction is sufficient to suggest that vicarious learning is occurring. See Online Appendix C for technical details.[4]

Vicarious Learning Given Volume of Experience. The effect of others' volume of experience is estimated as the average interfirm correlation between a firm's failure trajectory and another's failure trajectory for past observations if the other's volume exceeds a threshold of adverse events. Similar piecewise models were used in previous learning studies (Baum and Ingram 1998, Kim et al. 2009). These models captured the effects of preselected age ranges for the focal firm by splitting firm ages into segments. Here, we focus on others that exceed a predefined threshold as our respondents suggested that this is how they evaluated failure reports in the MAUDE data set. The threshold is iteratively increased in 10-report increments, from 0 to 600 reports per month, until $t-1$. For example, the effects for others with over 100 adverse events per month are obtained by calculating the average correlation between a firm's failure trajectory and that of another firm when adverse events for the other firm exceeded 100 per month. See Online Appendix C for additional details.

Vicarious Learning Given Delays. The effect of delays is calculated as the average interfirm correlation between a firm's failure trajectory and another's failure trajectory for those past observations when the other's delays were greater than a specified threshold. We do so in two steps. First, we measure the average delay for each other firm in each period, $D_{j,t}$. Second, we estimate the correlation of failure trajectories between the focal firm and another firm's trajectory when the other firm's past observations have delays that exceed a threshold. The threshold increases in 10-day increments, from 0 to 500 days. For example, the effect of delays greater than 50 days on vicarious learning is obtained by calculating the average correlation between the failure trajectory of a firm and another firm in those months when the delay of that other firm exceeds 50 days. See Online Appendix C for additional details.

Controls

In addition to controlling for vicarious learning through experience stock and referent effects, our analyses also accounted for firm-specific covariates: we controlled for each firm's ability to learn from its own experience, the timing for disclosing its own adverse events, and each firm's learning incentive and ability (via competition, diversification, and ownership effects). The models also included controls for

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

Table 1. Summary of Control Variables Used in the Regression

Controls	Variable symbol	Sample precedent	Mechanism to control	Operationalization
Industry avg. exp.	$IF_{j,t-1}$	Baum and Dahlin (2007), Greve (2003), Strang and Patterson (2014)	Flow of experience and social referent effects	Average number of adverse events reported by all other firm's at time $t-1$
Industry exp. stock	$IS_{j,t-1}$	Argote et al. (1990), Kim and Miner (2007), Madsen and Desai (2010)	Stock and forgetting of industry experience	Depreciated sum of all adverse event reports in the industry between 1997 to period $t-1$
Month	*MONTH*	Argote (2013), Thompson (2012)	Technological progress	Month time trend
No. of failures	$R_{i,t-1}$	Argote et al. (1990)	Path dependency	One-month lagged count of adverse events
Experience	$E_{i,t-1}$	Darr et al. (1995), Thompson (2007)	Learning and forgetting from firm experience	Log of the discounted sum of adverse events until $t-1$ for firm i
Disclosure	$d_{i,t-1}$	None	Disclosure of own events	Logged days between the occurrence and the reporting of each adverse event in month $t-1$
Diversification	$DIVERSE_{i,t-1}$	Joseph and Gaba (2015)	Scope economies and multimarket competition	Entropy measure the proportion of products a firm has in each medical sector
Competition	$COMPETITOR_{i,t-1}$	Baum and Ingram (1998)	Density-dependent competition effects	Firm's number of competitors in its main industry sector
Complexity	$COMPLEXITY_{i,t-1}$	Desai (2015), Haunschild and Sullivan (2002)	Failure complexity and distribution effects	Entropy measure of the proportion of adverse events in each medical sector observed by firm i
Publicly owned	$PUBLIC_{i,t-1}$	None	Reputation effects	Dummy variable that is equal to 1 if public firm, 0 if not
Size, firm	$SIZE_{i,t-1}$	Baum and Dahlin (2007), Madsen (2009)	Operational scale economies	Log of the number of products each firm i manufactured in past
Size, referent	$SIZE_{j,t-1}$	Baum and Dahlin (2007)	Others' operational scale economies	Logged average of the number of products for all other firms in $t-1$
New prod., firm	$NPI_{i,t-1}$	Maslach (2016), Srinivasan et al. (2007)	Risk of and ability to respond to problems	The number of a firm's new product introductions within 365 days preceding the start of period t
New prod., referent	$NPI_{j,t-1}$	None	Others' risk of and ability to respond to problems	The average number of new product introductions by other firms
Severity, firm	$SEVERITY_{i,t-1}$	Baum and Dahlin (2007)	Attention to salient events	The number of deaths reported by firm i
Severity, referent	$SEVERITY_{j,t-1}$	Madsen and Desai (2010)	Attention to others' salient events	The average number of deaths reported by other firms
Regulation	$REGULATION_{i,t-1}$	None	Regulatory compliance	Number of regulations experienced by each firm
MDUFMA	*MDUFMA*	Maslach (2016)	Regulatory changes	Dummy that indicates 1 after 2002, 0 otherwise

Notes. White indicates vicarious control; light gray indicates firm control, medium gray indicates failure control, and dark gray indicates regulatory control. MDUFMA, Medical Device User Fee and Modernization Act of 2002.

firm size, the severity of a failure, and new product introductions by each firm. Finally, we accounted for trends and changes in regulation over this period. While Table 1 presents all of the control variables we included in our analyses, for brevity, we highlight only the particularly relevant control variables in the main text: vicarious-specific experience, firm-specific experience, and disclosure. (Online Appendix D presents definitions of the remaining control variables.)

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

Vicarious-Specific Controls

Two measures of others' experience have been used in the past: (1) industry averaged experience (Baum and Dahlin 2007, Greve 2003, Strang and Patterson 2014) and (2) industry experience stock, which is measured as the discounted accumulation of all industry knowledge (Argote et al. 1990, Kim and Miner 2007, Madsen and Desai 2010). The former captures the social learning that occurs when firms benchmark against others, and the latter captures learning from a larger pool of knowledge.

Industry Averaged Experience. To control for the industry peer effects, we included the average number of adverse events reported by all other firm's at time $t-1$.

Industry Experience Stock. To control for industry stock, we included the depreciated sum of the number of all adverse event reports in the industry between 1997 and period $t-1$. We estimated the depreciation rate to be 0.99—a slow rate of industry forgetting—based on a grid search that maximized the explained variance of our estimates (Argote et al. 1990, Madsen and Desai 2010).

Firm-Specific Controls

Experience. Each firm had direct experience that afforded learning from its own past adverse events. Following prior research on learning from failures (Baum and Dahlin 2007, Kim and Miner 2007, Madsen and Desai 2010), we operationalized experience, $E_{i,t-1}$, as the log of the discounted sum of adverse events until $t-1$ for firm i. Following prior research on forgetting (Darr et al. 1995, Thompson 2007), we performed a grid search to identify the most likely discount parameter. Failure experiences decay relatively rapidly with a parameter of 0.75, indicating that they decay by half every 2.4 months. We also included a one-month lag for past failure to control for autocorrelation.

Disclosure. For each firm, we measured the average delay in disclosing its own adverse events in the prior month ($d_{i,t-1}$). We measured each delay as the logged number of days between the occurrence and the reporting of each adverse event in month $t-1$.

Analyses

We used a random effects negative binomial panel estimator (Wooldridge 2000) on the main sample. We chose the negative binomial over the Poisson model because we have significant overdispersion. The Hausman statistic for random versus fixed effect estimation was insignificant ($p < 0.49$), indicating that the random effect estimator was safe to use. The reported results are robust to alternative specifications, including random effects generalized least squares predicting the log rate of adverse events for each firm, the first difference in the number of adverse events reported by that firm from the previous year ($t-1$) to the current year (t), and the number of injuries for each firm. (These results are available upon request.)

Results

Table 2 presents abbreviated regression results for adverse events per firm. For brevity, the pairwise correlations and full results are reported in Online Appendix D. The log-likelihood of the model with just control variables is −11,367. Adverse events increased with a firm's own failure experience but decreased with longer delays in disclosing a firm's own failure reports. Model 2 shows that neither industry average experience nor industry experience stock had significant effects. If we stopped with these effects, it would appear that no vicarious learning was occurring in the industry. In the following discussion, we explore our alternative specification of vicarious learning.

Hypothesis 1 predicted that vicarious learning was occurring in the industry. For firms with historically similar patterns of adverse events, we predicted that an increase in the number of adverse event reports by that set of relevant others would lead to a decrease in subsequent adverse events in the focal firm. Model 3 supports our prediction: the coefficient is negative and significant ($p < 0.001$).

Hypothesis 2 predicts that firms will be less likely to learn as the volume of failure experience from others increases. As we increase the volume threshold from zero to the mean (comparing model 4 to model 3), the coefficient for vicarious learning given the volume of experience increases (i.e., from −0.82 to −0.40). Figure 1 shows how this coefficient changes as we iterate through volume and delay thresholds. There are two important points in interpreting these results. First, the volume and delay thresholds are for others' adverse event reports, not the firm's own reports. Second, as we iterate toward the tail of the distribution on volume or delay, the number of observations of others' diminishes in a smooth monotonic fashion, yet for each threshold, the smaller number of others' events has a greater effect. The effect appears to be nonmonotonic at volumes greater than 300 reports, and its significance changes as the upper bound of the confidence interval moves farther away from zero. This nonmonotonicity may reflect the moderating effect of delays hypothesized in H4.

We theorized in H3 that delays result in less vicarious learning. As we increase the reporting delay threshold from zero to the mean (comparing model 5 to model 3), the coefficient for vicarious learning given the volume of reporting delays increases (i.e., from −0.82 to −0.30), suggesting support for H3. Vicarious learning appears to decrease when others have longer reporting delays. As Figure 1 illustrates, the

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

Table 2. Abbreviated Results for Adverse Event Rates for Medical Device Firms, 1997–2012

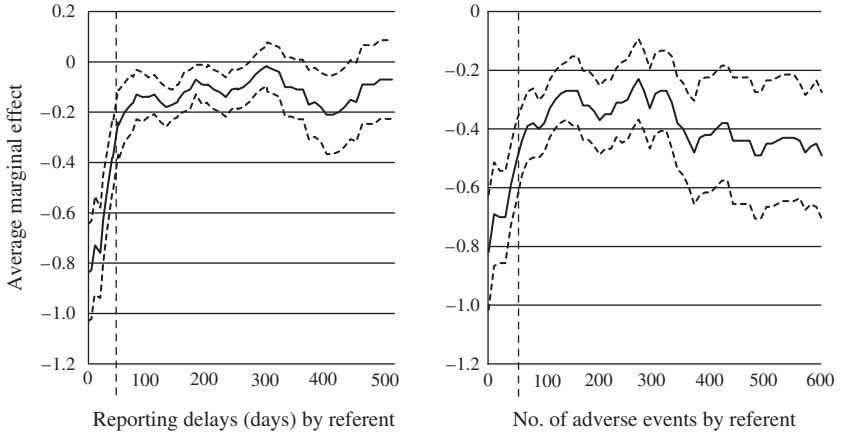
Variable name	Variable symbol	1	2	3	4	5	6	7
H1: Vicarious learning	$E_{j,t-1}$			−0.82***				
				(0.10)				
H2: Vicarious learning | Volume	$E_{j,t-1,>69\ \text{Reports}}$				−0.40***			
					(0.06)			
H3: Vicarious learning | Delays	$E_{j,t-1,>46\ \text{Days}}$					−0.30***		
						(0.07)		
H4: Vicarious learning | Volume, delays	$E_{j,t-1,>46\ \text{Days}>69\ \text{Reports}}$						−0.11*	
							(0.05)	
H4: Vicarious learning | Volume, delays	$E_{j,t-1,>46\ \text{Days}>200\ \text{Reports}}$							−0.13*
								(0.06)
Vicarious controls								
Industry avg. exp.	$IF_{j,t-1}$		−0.05	−0.04	−0.08+	−0.05	−0.06	−0.03
			(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)
Industry exp. stock	$IS_{j,t-1}$		0.03	−0.08	0.02	0.00	−0.10	0.18
			(0.13)	(0.05)	(0.13)	(0.13)	(0.16)	(0.23)
Relevant firm controls								
No. of failures	$R_{i,t-1}$	0.05	0.04	0.03	0.02	0.03	0.02	−0.02
		(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)
Experience	$E_{i,t-1}$	0.57***	0.57***	0.64***	0.63***	0.58***	0.59***	0.67***
		(0.05)	(0.05)	(0.05)	(0.05)	(0.03)	(0.05)	(0.05)
Disclosure	$d_{i,t-1}$	−0.04***	−0.04***	−0.07***	−0.05***	−0.06***	−0.04*	−0.02+
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Log-likelihood		−11,367	−11,366	−11,331	−11,274	−11,351	−10,311	−7,893
Wald test			0.52	71.6***	50.1***	20.4***	5.3*	4.7*

Notes. See Online Appendix D for the full table of results with additional controls. We include firm-level controls for time trend, diversification, competition, complexity, and public ownership. We also include variables to control for failure-related effects. These variables include the size of the firm and size of the referent, new product introductions for the firm and the referent, and failure severity for the firm and referent. Finally, we control for the number of regulations and the MDUFMA—a major policy change in the industry. N = 2,170 firm-month observations. N = 1,951 for model 6. N = 1,445 for model 7. The Wald test is based on the "test" postestimation command on the variable of interest in Stata, where we tested whether the coefficient differed from zero.

***, *, and + represent $p < 0.001$, $p < 0.05$, and $p < 0.1$, respectively.

significance and negative magnitude of the coefficient diminishes with high thresholds of others' reporting delays. After one year, the effect of vicarious learning is only 18% of the effect size for delays reported within the first 30 days. There is a small spike shortly after 10 days because the FDA permits a 10-day grace period for reports. The effect of delays on vicarious learning is uniformly negative.

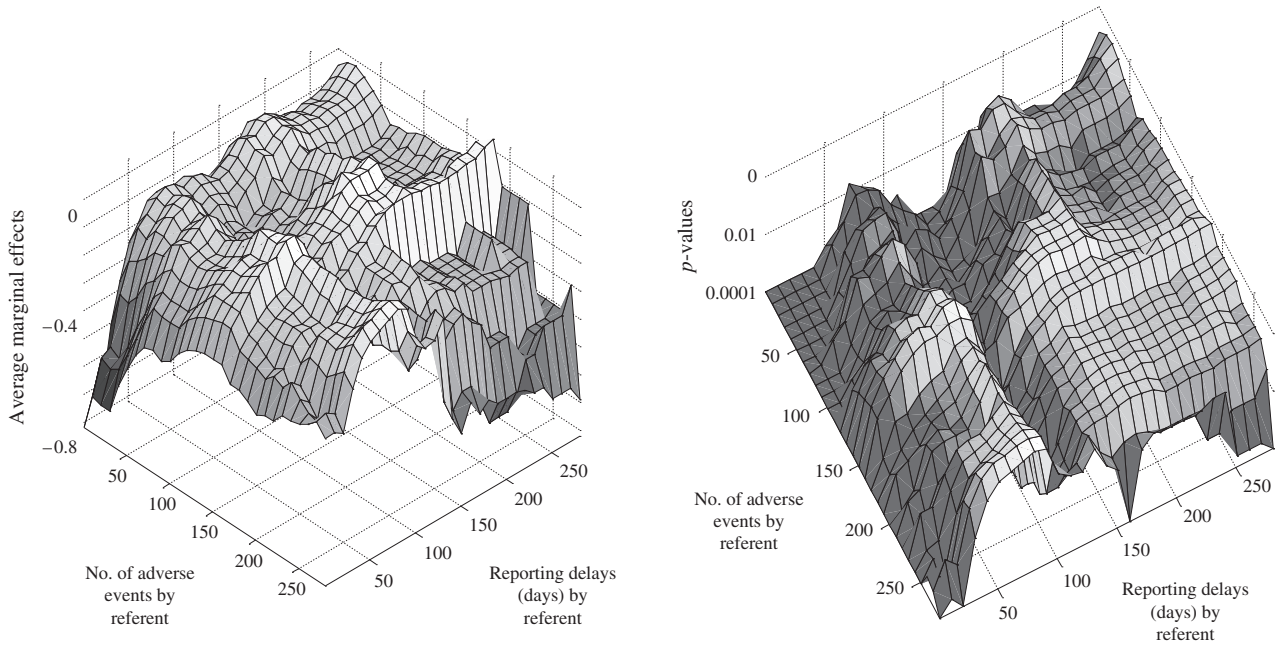
Figure 1. Effects of Vicarious Learning Given Changes in Reporting Delays and Volume of Experience



Notes. The vertical lines indicate the average logged delay (46 days) and average logged volume (69 reports), respectively. Dashed lines are 95% confidence intervals.

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 2. The Effects of Others' Volume and Delays on Vicarious Learning



Hypothesis 4 predicted that delays may signal learning opportunities when failure reports are scarce and adverse event reports take a long time to appear or are being reclassified. Hypothesis 4 receives some support by comparing delayed experiences at the mean volume to high volumes in models 6 and 7. The omnibus model significance increases and the effect size of the coefficient increases modestly, but the Stata "test" postestimation command suggests that its significance does not change. However, such a simple comparison is limited because the number of observations of others' failures changes as the delay threshold changes. The three-dimensional graphs in Figure 2 provide a more complete picture of the nonlinear effects on learning. Three effects stand out. First, the effects follow an inverted-U shape centered on reporting delays around 80 days. For delays shorter than 80 days, each additional day of delay leads to less vicarious learning. For delays longer than 80 days, each additional day of delay leads to more vicarious learning. Second, as hypothesized in H4, the dampening effect of delays on vicarious learning appears to be more severe when the volume of experience is low rather than high.[5] The inverted-U shape exists at both low and high volumes, but the lowest point of the inverted U (and hence the largest effect on learning) can be found at volumes below 100 for others' adverse event reports. (See the dark shaded area on the leftmost part of the left graph in Figure 2.) Third, the significance and effect size increase for extreme volumes and delays. In particular, both the effect size and the significance of the learning are greatest for referents with 250 reports that are more than 250 days late. (See the dark shaded areas on the right in both graphs in Figure 2.)

Post Hoc Tests

Our results remained substantively unchanged across several robustness checks. (Additional details can be found in Online Appendix B.) Notably, the results are robust for all three sampling strategies: the entire industry, firms with more than one adverse event, and the sample reported here. They held when we modeled the nonreported experiences as zeros, but multicollinearity was high. They also held when we substituted injury reports rather than just reports of adverse events, or when we added additional control variables for each firm's age and country of origin, suggesting that the model is robust to different specifications. We also controlled for selection effects through Heckman selection models and multiple techniques (i.e., Cox proportional hazards survival analysis) to show that the effects are comparable.

Text Analysis. To explore why some reports were delayed, we content analyzed the narrative text in 6,769 adverse event reports in 2007. (Results are available upon request.)[6] We wanted to measure the comprehensiveness and the clarity of the text to see whether the text improved with delays. Our negative binomial model regressed the number of days a report was delayed on four variables. Text comprehensiveness was measured by two variables: (i) the logged number of characters in the manufacturer narrative explaining the root causes of an adverse event and (ii) the logged number of characters in the event description. On the

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

basis of the assumption that a narrative with fewer words is likely to be clearer than a narrative with more words (Flesch 1948), we included two additional variables that focused on clarity: (i) the logged word count of the manufacturer narrative that explained the root causes of the adverse event and (ii) the logged word count of the event description.

Reports with longer delays had text that was more comprehensive. Delayed reports had more text characters ($p < 0.001$) in the event descriptions and the narratives of why the events occurred. Delayed reports were also clearer. They had fewer words in the event descriptions ($p < 0.001$), although the coefficients for word count of the narrative of the event's root causes were insignificant. Delayed reports provided descriptions of events that were both more clear and comprehensive, suggesting that they provide better information about the failure.

Discussion

> The pursuit of rich experience, however, requires a method for absorbing detail without molding it. Great organizational histories, like great novels, are written, not by first constructing interpretations of events and then filling in the details, but by first identifying the details and allowing the interpretations to emerge from them. As a result, openness to a variety of (possibly irrelevant) dimensions of experience and preference is often more valuable than a clear prior model and unambiguous objectives. (March et al. 1991, p. 8)

A repository such as the MAUDE data set accumulates the adverse events of an industry in the hopes of learning from what is usually considered noise. Unlike typical learning models, a repository of failure events provides no clear signal of success to imitate. It publicizes the experiences of a variety of individuals and organizations but only in the form of failure reports. It would be tempting to follow existing approaches to look for a signal in this noise, but doing so misses the point of the repository.

Repository-based learning combines both valid and reliable learning processes. Our qualitative evidence showed valid learning processes as actors sampled adverse events by looking for evidence of either failure modes or protocol violations. They then used the details of those adverse events to identify patterns that led to new understandings around device failure, thereby generating valid learning processes. Our quantitative evidence sought to show reliable learning processes as actors used the volume of failures to find a signal around a specific protocol violation, reporting delays.

In this combination of learning processes, repository-based learning challenges traditional understandings of vicarious learning from failures. First, in contrast to learning from referent others (Kacperczyk et al. 2014), which draws a clear signal from success, and learning from rare events, which draws a signal from the salience of an event (Lampel et al. 2009), repository-based learning draws a signal from the noise. Second, in contrast to learning from referent others, which relies on aggregate evidence of success and failure, the repository requires that firms be able to learn from the details of specific events. Third, in contrast to learning from referent others, in which the context is determined by the reference set, and learning from rare events, in which the context is established by the specific event, repository-based learning allows firms to build the context around experience. These contrasts reveal a different form of vicarious learning but also present significant empirical challenges. Our findings explore some possibilities for overcoming those challenges and suggest directions for thinking about an increasingly important form of data for vicarious learning.

Noise as Signal

Theories of vicarious learning typically assume that firms copy the experience of more successful firms in their competitive set (Greve and Taylor 2000, Haunschild and Miner 1997, Srinivasan et al. 2007). These findings describe a reliable learning process in which firms should copy the signal contained within success. Conversely, firms should ignore referent others that are less successful, because their failures demonstrate that they do not know what they are doing. Standard learning models rely on signals of probability, assuming that the known experiences of the most effective competitors are most likely to transfer successfully to their own experience. In general, this is wise because success is a reliable signal that learning by imitation will be fruitful (Haunschild and Miner 1997, Kacperczyk et al. 2014).

A repository of failure events presents a challenge to this assumption, because the signal in the repository must emerge from what we usually treat as noise. The MAUDE data set asks firms to use the failure data as signals of possibility. The repository displays failure experiences as hypothetical histories—things that did not happen to the firm but *could* have happened to them (March 1994). For instance, in the example of the computer crashing during surgery, the MAUDE data helped other firms see how it would be possible for any computer-aided device to crash in a similar manner. In other instances, the repository made details and interpretations of a similar experience available, such as when our informants used evidence of burns from a cosmetic device to understand aspects of the device use that they did not expect. In each of these instances, the signal in the repository lies in both the possibility of the failure and in the associated details of that failure. This combination opens up multiple realms of possibility that the firm may not have anticipated.

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

The signal in a repository can also change across time and space. In typical vicarious learning, the learning is episodic: once the best firms learn from their experience, the other firms immediately copy that experience. Those episodes are also constrained in space because the learning occurs between the set of firms in a competitive referent set. Learning from a repository emerges from events that can be selected across both space and time, yielding a signal that can change as events accumulate. As described in the quote that opens our discussion, the learning begins with details that can then be used to construct interpretations (March et al. 1991). In searching the repository, actors can assemble and reassemble failure events that may have occurred at multiple points in time or from multiple different actors. Each failure event added to the repository adds new details and new lessons. Learning from the noise depends on how actors sort through the events and absorb the new details. We turn to that learning process next.

Learning from Noise

The research on organizational learning has followed a progression from focusing on evidence of learning from aggregate total prior experience (Argote and Epple 1990, Darr et al. 1995, Ingram and Baum 1997) to disaggregating failure experience and near-failure experience from total prior experience (Dahlin et al. 2018, Desai et al. 2018, Gong et al. 2017, Kim and Miner 2007, Kim et al. 2009, Sitkin 1992). For example, Haunschild and Sullivan (2002) show that heterogeneity of a firm's own failure experience can lead to reduced accident rates through a deeper and broader search for causes. More recent research has compared the effectiveness of vicarious learning from success and failure (e.g., Madsen and Desai 2010) and shown whether firms learn from the failures of others (Madsen 2009). This research shows that success and failure make different contributions to learning, but they can be used systematically (Beckman and Haunschild 2002) to increase the signal and diminish the noise of failure (Bohn 1995).

Our study of learning from a repository of failure events extends this progression by showing how individuals and firms learn from the *specific* failure experiences of others. Learning from rare events comes close to such learning from specific events, but that learning depends on the arrival of a singular and salient rare event. By contrast, learning from a repository is intentional: firms actively choose to search the repository to find events from which they might learn.

This repository-based inferential learning process is a form of problemistic search—search stimulated by a problem (Argote and Greve 2007, Cyert and March 1963, Posen et al. 2018, Vissa et al. 2010)—but with four variants. First, in problemistic search, the failure to meet aspiration levels triggers search (Gavetti et al. 2012, Greve 2003, Ref and Shapira 2017), so the attention is driven by a firm's own experience, its aspiration levels, and the associated comparisons to others. By contrast, repository-based learning may be driven by a firm's own experience, as when a firm wants to compare its adverse events to those of others, but the firm may also choose to search the repository for new experience. Both the qualitative evidence of search strategies (the failure modes and protocol violations) and the quantitative evidence (volume and delays) point to active and intentional search strategies.

Second, repository-based problemistic search is often prospective. Learning research typically focuses on how firms evaluate past experience against aspiration levels (Shinkle 2012). As Lant and Shapira (2008, p. 60) observe, "A key assumption of the behavioral theory of the firm is that firms adjust their behavior in response to their experience rather than acting on their expectations of future states of the world." There is evidence of retrospective search in our data, but much of our qualitative and quantitative evidence focuses on instances where individuals actively query the MAUDE database to anticipate future problems. For instance, failure mode searches are used to help organizations imagine possible problems so that the device design and instructions can prevent them before they ever happen.

Third, typical problemistic search depends on the availability of slack resources (Cyert and March 1963, Posen et al. 2018, Ref and Shapira 2017, Vissa et al. 2010). The repository substantially reduces the resources required for search. In cases such as the computer device that shut down in the midst of the procedure, firms do not even need to engage in search, because the problem is made publicly available. In other cases, the MAUDE data set extends the available range of events, so the firm is not limited to problemistic search in the neighborhood of existing practice (Baum and Dahlin 2007) because it has ready access to the experiences of an entire industry.

Finally, time is a factor. Typically, problemistic search is stimulated by the firm's failures at a particular point in time, so the outcome of the search is constrained by the practices of other firms at that moment in time. By contrast, a repository makes available a range of experiences across time. Firms can search adverse events across that entire range as possible stimuli for problems, solutions, or both.

Our evidence on repositories describes an inferential learning process through which actors learn by "absorbing detail without molding it" (March et al. 1991, p. 8). Actors select adverse events from the repository and then pool features of an event—different aspects, different interpretations, and different preferences—and assemble those features into

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

a new understanding. For example, our informants sometimes used the MAUDE data set to see how people were using their devices in unexpected ways. In other instances, rather than beginning with a reference set, firms collected adverse event reports and then constructed the reference set. These moments of learning were turning points in history, but unlike with rare events, where the turning point might be driven by the salience of the event, in these cases the turning point was constructed through the learning process. The learning happens through a shift in belief, signaled by the new possibilities that emerge from what standard models of learning might treat as noise.

Constructing Context

A repository of failure allows firms to shape the context for their learning (cf. Argote and Greve 2007, p. 342) by choosing a broader range of referent others. When firms draw their lessons from successful others, they may undersample failure and draw biased conclusions (Denrell 2003). The repository redresses that problem by offering an excess of failures. Although the probability estimates remain biased, the repository exposes firms to an array of detail from which they can build new interpretations (March 2010, March et al. 1991). This exposure is especially important to the firms that do not have frequent failures, because they may be unaware of both potential failure modes and the details of the associated events.

This form of learning suggests that there are public good aspects of failure (Knott and Posen 2005). Rather than thinking about failures as something to be suppressed, the idea of a repository presumes that there is merit in sharing adverse events. This introduces a tension in how we think about failure. On the one hand, more successful firms understandably want to demonstrate superiority by having fewer failures, but better performance also reduces the set of events from which others can learn. On the other hand, if we encourage firms to report, the aggregate outcome can look like a failure to learn because the rate of adverse events will appear to increase. There is a subtle but important distinction that needs to be considered here. For a repository to be effective, firms must embrace the reporting process by reporting events that they would prefer to keep hidden. Firms may also report near failures and less significant events (Madsen et al. 2016), because they want to expand the range of events available for vicarious learning. As a result, the number of failure events may therefore increase over time, as it does in our data.

Contextual details are also important to how we think about big data. A repository of failure events seeks to stimulate vicarious learning by making data publicly available (Tamuz 2001, Thaler 2011). Most approaches to big data treat events as independent samples from a universe and then rely on algorithms to sort out signal from noise (George et al. 2016). With a repository, the individuals are free to creatively select and combine adverse events and contextual details that they consider most applicable to their products, so they may identify problems that would be impossible to detect with preestablished, rule-based algorithms. Such learning is important to medical devices, but it also extends to everyday contexts such as the safety of consumer products (Thaler 2011) and child care products (Eig 1998). Understanding the benefits of such data can aid the government in regulatory actions because companies sometimes worry that making failure data publicly available is harmful. Our findings invert those concerns and show how firms learn from the failures of others and improve their own products.

Limitations

Following calls for research on the processes of vicarious learning from failure (e.g., Baum and Dahlin 2007, p. 382), we used the adverse event reports in the MAUDE data to show more direct evidence of learning. In our qualitative evidence, we showed how individuals intentionally chose adverse events from firms outside their competitive set to identify possibilities that they had not experienced or imagined. This is not a statistical form of learning, but rather one in which new, valid insights emerge from the details of individual events (March et al. 1991, p. 8). New insights can then lead to process or product changes that in turn lead to fewer failure events in the future. Our quantitative findings therefore focus on failure events, not failure rates, to demonstrate learning.

This approach should get us closer to the actual learning process, but it also presents some analytic challenges. Our quantitative data may look like weak evidence of learning because they do not show clear performance improvement. The MAUDE data show an increase in adverse event reports across time, which does not look like aggregate learning. If decreased failure rates are our only measure of vicarious learning, then it is possible that no genuine learning occurred. The increase in reported events may reflect firms learning to comply with regulations. That is still a form of learning, but it does not show improved performance. It is, however, consistent with evidence showing that a positive safety climate leads to increased reporting of near misses (Dillon et al. 2016). It is also consistent with the views of our respondents. Our informants said that they valued the MAUDE data set for the lessons that they drew from the descriptions of others' adverse events. They also said they willingly contributed reports of their own adverse events. If their claims are accurate, then the increase in adverse event reports may reflect firms' improved understanding of and commitment to the MAUDE data as a source of vicarious learning.

Published in Organization Science on April 02, 2018 as DOI: 10.1287/orsc.2017.1179.
This article has not been copyedited or formatted. The final version may differ from this version.

The report counts we use to establish reporting patterns also have some analytical limitations. The MAUDE data draw on accounts of varying quality, following diverse reporting practices, about products from a very heterogeneous set of firms. Those sources provide diverse information, but it is hard to both capture that diversity and make effective comparisons without failure rates and clear referent others. Medical device failures can have serious and significant consequences, so adverse events should be infrequent. Yet, without a baseline of failure events, it is impossible to show that failure rates decrease with experience. Such problems are not unique to our research and may be endemic to studying learning. Standard models of learning show a relationship between cumulative experience and improvement in measures such as cost and quality, but establishing the causes of those improvements is challenging (Adler and Clark 1991, Sinclair et al. 2000, Thompson 2012). More recent approaches to learning have developed more fine-grained characterizations of experience, context, and knowledge transfer to resolve such problems (Argote and Miron-Spektor 2011).

In that spirit, we focus on learning from failure, but demonstrating learning from failure presents other challenges. Standard econometric methods may produce erroneous results in demonstrating learning from failure (Bennett and Snyder 2017), so here, we demonstrate learning by comparing failure events across firms that report consistently over time. The correlations that we report could be driven by learning, but they could also be driven by similarities between firms or by exogenous events that affect all firms. The correlations may also understate the evidence of learning, because they would miss instances in which only a subset of our sample of firms learn from each other. We caution about the generalizability of our findings because we cannot show the significance of these effects.

Our methods also present some interpretive challenges. To demonstrate learning, we compare coefficient size and significance as we iterate across changes in volume and delays in the reports of other firms. Our method is similar to more recent approaches used to interpret the interaction effects in nonlinear models (Zelner 2009). These approaches demonstrate the effects of discrete changes in the values of independent variables (over a range of these variables) on the predicted probabilities of the outcome variable. Graphical techniques are advantageous because one can more easily observe a relationship over a landscape of values of independent variables, particularly if the relationships are difficult to model more directly. Our graphical depiction shows the effect of past interfirm correlations on future failure and how this effect varies with changes in the values of others' volume and delay. It is difficult to provide linear estimates in our model as in past research, so we cannot compare the differences between points in our graphical representation, but that was not our goal. Instead, we wanted to follow the users of the MAUDE data to identify where the combinations of others' volume and delay were most likely to effect learning.

Although our quantitative methods may not meet the standards of existing research for demonstrating reliable learning processes, they do suggest directions for further research. The problems we encountered could be addressed with even more fine-grained data that directly match the flow of adverse event accounts across firms. For instance, a more homogeneous repository of failure events could be used to track how the text of failure reports draws from the lessons of others across different categories of failure events. Furthermore, with a repository of failures from a less heterogeneous set of firms and products, it might be possible to use failure rates to demonstrate the effects of specific adverse events on subsequent performance. Finally, it might be possible to select on severe adverse events and again construct categories of failure or focus on the effects of those severe failures on subsequent failure rates. These approaches would move future research closer to the combination of rich learning and reinforcement learning that we set out to explore in the MAUDE data set.

Conclusion

Our study focuses on learning from specific failure reports of medical device firms. Improvements in computing power, data storage, and communication have substantially reduced the costs of producing and maintaining a data set such as the MAUDE repository, suggesting that such repositories are likely only to increase in prevalence. Regulators such as the FDA can make available participant-produced data to improve safety with only very limited oversight or direct intervention (Thaler 2011). The merits of such data are clear. Users have access to the experience of a broad range of others in their industry, and in turn, researchers can get more directly at the process by which firms develop knowledge. Rather than rely on units of production and failure rates as a proxy for knowledge, we can turn directly to reports of specific events. With others, we hope to chart a direction toward learning from product and problem diversity both within (Egelman et al. 2017) and across (Nickerson et al. 2017) individuals and organizations.

Following that direction requires a move beyond some standard assumptions about learning from the failures of others. One departure is that firms learn from referent others. We show that in addition to learning from the best practices of others, firms can also use repositories to learn directly from the failure events of others. Understanding this form of learning requires

that scholars find a signal in what would ordinarily be treated as noise. Our qualitative exploration shows that such learning is possible. Our quantitative methods show how one might provide empirical support of such learning. The challenge for future research is showing more directly how that signal leads to learning from the failures of others.

Acknowledgments

The authors are grateful to senior editor Zur Shapira and three anonymous reviewers who went beyond the call of duty in clarifying their message. They also thank Joel Baum, Mark Bergen, You-Ta Chuang, Kristina Dahlin, Natalie Bin Zhao, Jerker Denrell, Vinit Desai, Dan Levinthal, Daniel Levy, Chengwei Liu, Peter Madsen, Fernando Olivera, Brian Pinkham, Stephen Sapp, Simon Parker, Hart Posen, Hazhir Rahmandad, Glenn Rowe, and Michal Tamuz, whose excellent comments improved this manuscript. They thank seminar participants at Oregon State University, University College London, the 2009 Organization Science Winter Conference, the 2009 and 2016 annual meetings of the Academy of Management, and the 2010 Medici Summer School of Management Studies for helpful comments on earlier versions of this manuscript. The authors are grateful for financial support from the Conrad Business, Entrepreneurship and Technology Centre at the University of Waterloo; Florida State University; the Social Sciences and Humanities Research Council Doctoral Fellowship; the Ontario Graduate Scholarship; the C. B. Johnston Ontario Graduate Scholarship; the R. G. Barford Ontario Graduate Scholarship; the Connor, Clark & Lunn Ontario Graduate Scholarship; and Western University. All remaining errors are the authors', and any findings and statements in this paper reflect the views of the authors, and not those of any other entity.

Endnotes

[1] Examples abound: for consumer product safety, see http://www.saferproducts.gov; for computer device failure, see https://www.usenix.org/cfdr; for disaster and failure events, see http://wtcdata.nist.gov; for medical devices, see https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm; for pharmaceutical products, see http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/; and for automobile accidents, see http://www.nhtsa.gov/NCSA.

[2] Prior to 1997, reports were voluntary. We dropped records from December 2012 to prevent truncated missing reports.

[3] To get to this sample, we excluded all 17,854 observations from firms that did not report in every month. We also removed 1,229 observations from which delays could not be calculated because dates were unreported. We lagged independent variables by one month.

[4] We tried other operationalizations such as measuring vicarious learning based on an aggregate count of others' failures in a period and transforming the data set to compare each firm to all other firms in dyadic relationships. These results were substantively similar to those using the current methodology but are more difficult to interpret.

[5] Results beyond 280 days for more than 280 adverse events were not obtained because the regressions did not converge.

[6] We limited our analysis to reports in one year because of computational difficulties in analyzing large text files. There were 1,679,921 text characters in our sample. We chose 2007 because it was near the midpoint of our sample. In the analysis, we controlled for the logged number of adverse events per firm and logged number of deaths per firm.

References

Abrahamson E (1991) Managerial fads and fashions: The diffusion and rejection of innovations. *Acad. Management Rev.* 16(3):586–612.

Adler PS, Clark KB (1991) Behind the learning curve: A sketch of the learning process. *Management Sci.* 37(3):267–281.

Argote L (2013) *Organizational Learning: Creating, Retaining and Transferring Knowledge* (Springer, Boston).

Argote L, Epple D (1990) Learning curves in manufacturing. *Science* 247(4945):920–925.

Argote L, Greve H (2007) *A Behavioral Theory of the Firm*—40 years and counting: Introduction and impact. *Organ. Sci.* 18(3):337–349.

Argote L, Miron-Spektor E (2011) Organizational learning: From experience to knowledge. *Organ. Sci.* 22(5):1123–1137.

Argote L, Beckman SL, Epple D (1990) The persistence and transfer of learning in industrial settings. *Management Sci.* 36(2):140–154.

Baum JAC, Dahlin KB (2007) Aspiration performance and railroads' rates of experiential and vicarious learning from train wrecks and crashes. *Organ. Sci.* 18(3):368–385.

Baum JAC, Ingram P (1998) Survival-enhancing learning in the Manhattan hotel industry, 1898–1980. *Management Sci.* 44(7):996–1016.

Beckman CM, Haunschild PR (2002) Network learning: The effects of partners' heterogeneity of experience on corporate acquisitions. *Admin. Sci. Quart.* 47(1):92–124.

Bennett VM, Snyder J (2017) The empirics of learning from failure. *Strategy Sci.* 2(1):1–12.

Bohn RE (1995) Noise and learning in semiconductor manufacturing. *Management Sci.* 41(1):31–42.

Chuang Y-T, Baum JAC (2003) It's all in the name: Failure-induced learning by multiunit chains. *Admin. Sci. Quart.* 48(1):33–59.

Cimpanu C (2016) Medical equipment crashes during heart procedure because of antivirus scan. *Softpedia* (May 3), http://news.softpedia.com/news/medical-equipment-crashes-during-heart-procedure-because-of-antivirus-scan-503642.shtml-ixzz4KKqyP7Bj.

Creswell JW (2003) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 2nd ed. (Sage, Thousand Oaks, CA).

Cyert RM, March JG (1963) *A Behavioral Theory of the Firm* (Blackwell Publishers, Malden, MA).

Dahlin KB, Chuang YT, Roulet TJ (2018) Opportunity, motivation and ability to learn from failures and errors: Review, synthesis and ways to move forward. *Acad. Management Ann.* 12(1):252–277.

Darr ED, Argote L, Epple D (1995) The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management Sci.* 41(11):1750–1762.

de Holan PM, Phillips N (2004) Remembrance of things past? The dynamics of organizational forgetting. *Management Sci.* 50(11):1603–1613.

Delporte C (2011) Top of the news: Guidant fined, given probation for defective heart devices. *Medical Product Outsourcing* 9(1):14–16.

Denrell J (2003) Vicarious learning, undersampling of failure, and the myths of management. *Organ. Sci.* 14(3):227–243.

Denrell J (2008) Superstitious behaviour as a byproduct of intelligent adaptation. Hodgkinson GP, Starbuck WH, eds. *The Oxford Handbook of Organizational Decision Making* (Oxford University Press, Oxford, UK), 271–286.

Denrell J, Le Mens G (2016) Information sampling, belief synchronization and collective illusions. *Management Sci.* 63(2):528–547.

Denrell J, Fang C, Levinthal DA (2004) From T-mazes to labyrinths: Learning from model-based feedback. *Management Sci.* 50(10):1366–1378.

Desai V (2015) Learning through the distribution of failures within an organization: Evidence from heart bypass surgery performance. *Acad. Management J.* 58(4):1032–1050.

Desai VM, Maslach D, Madsen P (2018) Organizational learning from failure: Present theory and future inquiries. Argote L, Levine JM, eds. *Oxford Handbook of Group and Organizational Learning*. Forthcoming.

Diehl E, Sterman JD (1995) Effects of feedback complexity on dynamic decision making. *Organ. Behav. Human Decision Processes* 62(2):198–215.

Dillon RL, Tinsley CH, Madsen PM, Rogers EM (2016) Organizational correctives for improving recognition of near-miss events. *J. Management* 42(3) 671–697.

DiMaggio PJ, Powell WW (1983) The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *Amer. Sociol. Rev.* 48(2):147–160.

Diwas K, Staats BR, Gino F (2013) Learning from my success and from others' failure: Evidence from minimally invasive cardiac surgery. *Management Sci.* 59(11):2435–2449.

Dye KC, Eggers JP, Shapira Z (2014) Trade-offs in a tempest: Stakeholder influence on hurricane evacuation decisions. *Organ. Sci.* 25(4):1009–1025.

Egelman CD, Epple D, Argote L, Fuchs E (2017) Learning by doing in multiproduct manufacturing: Variety, customizations, and overlapping product generations. *Management Sci.* 63(2):405–423.

Ehrenkranz M (2016) Heart machine crashes mid-procedure because it was busy checking itself for viruses. *Mic* (May 9), https://mic.com/articles/142989/heart-machine-crashes-mid-procedure-because-it-was-busy-checking-itself-for-viruses-.yB80QitjP.

Eig J (1998) How Danny died. *Chicago Magazine* (November), 62–82.

Flesch R (1948) A new readability yardstick. *J. Appl. Psych.* 32(3):221–233.

Garud R, Dunbar RLM, Bartel CA (2011) Dealing with unusual experiences: A narrative perspective on organizational learning. *Organ. Sci.* 22(3):587–601.

Gavetti G, Greve H, Levinthal D (2012) The behavioral theory of the firm: Assessment and prospects. *Acad. Management Ann.* 6(1):1–40.

Geddes LA (1995) *Handbook of Electrical Hazards and Accidents* (CRC Press, Boca Raton, FL).

George G, Osinga EC, Lavie D, Scott BA (2016) Big data and data science methods for management research. *Acad. Management J.* 59(6):1493–1507.

Gibson CB (2017) Elaboration, generalization, triangulation, and interpretation: On enhancing the mixed method research. *Organ. Res. Methods* 20(2):193–223.

Gong Y, Zhang Y, Xia J (2017) Do firms learn more from small or big successes and failures? A test of the outcome-based feedback learning perspective. *J. Management*, ePub ahead of print January 30, https://doi.org/10.1177/0149206316687641.

Greve HR (2003) *Organizational Learning from Performance Feedback: A Behavioural Perspective on Innovation and Change* (Cambridge University Press, Cambridge, UK).

Greve H, Taylor A (2000) Innovations as catalysts for organizational change: Shifts in organizational cognition and search. *Admin. Sci. Quart.* 45(1):54–80.

Haunschild PR, Miner AS (1997) Modes of interorganizational imitation: The effects of outcome salience and uncertainty. *Admin. Sci. Quart.* 42(3):472–500.

Haunschild PR, Sullivan BN (2002) Learning from complexity: Effects of prior accidents and incidents on airlines' learning. *Admin. Sci. Quart.* 47(4):609–643.

Hoffman AJ, Ocasio W (2001) Not all events are attended equally: Toward a middle-range theory of industry attention to external events. *Organ. Sci.* 12(4):414–434.

Ingram P, Baum JAC (1997) Opportunity and constraint: Organizations' learning from the operating and competitive experience of industries. *Strategic Management J.* 18(S1):75–98.

Joseph J, Gaba V (2015) The fog of feedback: Ambiguity and firm responses to multiple aspiration levels. *Strategic Management J.* 36(13):1960–1978.

Kacperczyk A, Beckman CM, Moliterno TP (2014) Disentangling risk and change: Internal and external social comparison in the mutual fund industry. *Admin. Sci. Quart.* 60(2):228–262.

Kim J-YJ, Miner AS (2007) Vicarious learning from the failures and near-failures of others: Evidence from the U.S. commercial banking industry. *Acad. Management J.* 50(3):687–714.

Kim J-Y, Kim J-YJ, Miner AS (2009) Organizational learning from extreme performance experience: The impact of success and recovery experience. *Organ. Sci.* 20(6):958–978.

Knott AM, Posen HE (2005) Is failure good? *Strategic Management J.* 26(7):617–641.

Kunreuther H, Bowman EH (1997) A dynamic model of organizational decision making: Chemco revisited six years after Bhopal. *Organ. Sci.* 8(4):404–413.

Labianca G, Fairbank JF, Andrevski G, Parzen M (2009) Striving toward the future: Aspiration-performance discrepancies and planned organizational change. *Strategic Organ.* 7(4):433–466.

Labib A, Read M (2013) Not just rearranging the deckchairs on the *Titanic*: Learning from failures through. *Risk Reliability Anal.* 51(1):397–413.

Lampel J, Shamsie J, Shapira Z (2009) Experiencing the improbable: Rare events and organizational learning. *Organ. Sci.* 20(5):835–845.

Lant TK, Shapira Z (2008) Managerial reasoning about aspirations and expectations. *J. Econom. Behav. Organ.* 66(1):60–73.

Lapré MA, Tsikriktsis N (2006) Organizational learning curves for customer dissatisfication: Heterogeniety across airlines. *Management Sci.* 52(3):352–366.

Levitt B, March JG (1988) Organizational learning. *Annual Rev. Sociol.* 4:319–343.

Li Q, Maggitti PG, Smith KG, Tesluk PE, Katila R (2013) Top management attention to innovation: The role of search selection and intensity in new product introductions. *Acad. Management J.* 56(3):893–916.

Lieberman MB, Asaba S (2006) Why do firms imitate each other? *Acad. Management Rev.* 31(2):366–385.

Madsen PM (2009) These lives will not be lost in vain: Organizational learning from disaster in U.S. coal mining. *Organ. Sci.* 20(5):861–875.

Madsen PM, Desai V (2010) Failing to learn? The effects of failure and success on organizational learning in the global orbital launch vehicle industry. *Acad. Management J.* 53(3):451–476.

Madsen PM, Dillon RL, Tinsley CH (2016) Airline safety improvement through experience with near-misses: A cautionary tale. *Risk Anal.* 36(5):1054–1066.

March JG (1994) *A Primer on Decision Making: How Decisions Happen* (Free Press, New York).

March JG (2010) *The Ambiguities of Experience* (Cornell University Press, Ithaca, NY).

March JG, Shapira Z (1992) Variable risk preferences and the focus of attention. *Psych. Rev.* 99(1):172–183.

March JG, Sproull LS, Tamuz M (1991) Learning from samples of one or fewer. *Organ. Sci.* 2(1):1–13.

Maslach D (2016) Change and persistence with failed technological innovation. *Strategic Management J.* 37(4):714–723.

Massini S, Lewin AY, Greve H (2005) Innovators and imitators: Organizational reference groups and adoption of organizational routines. *Res. Policy* 34(10):1550–1569.

Miner AS, Kim J-Y, Holzinger IW, Haunschild P (1999) Fruits of failure: Organizational failure and population level learning. Miner AS, Anderson P, eds. *Population-Level Learning and Industry Change*, Advances in Strategic Management, Vol. 16 (JAI Press, Greenwich, CT), 187–220.

Moliterno TP, Beck N, Beckman CM (2014) Knowing your place: Social performance feedback in good times and bad times. *Organ. Sci.* 25(6):1684–1702.

Nickerson JA, Wuebker R, Zenger T (2017) Problems, theories, and governing the crowd. *Strategic Organ.* 15(2):275–288.

Ocasio W (1997) Towards an attention-based view of the firm. *Strategic Management J.* 18(S1):187–206.

Ocasio W (2011) Attention to attention. *Organ. Sci.* 22(5):1286–1296.

Pisano GP, Bohmer RMJ, Edmondson AC (2001) Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Sci.* 47(6):752–768.

Porac JF, Thomas H, Baden-Fuller C (1989) Competitive groups as cognitive communities: The case of Scottish knitwear manufacturers. *J. Management Stud.* 26(4):397–416.

Porac JF, Thomas H, Wilson F, Paton D, Kanfer A (1995) Rivalry and the industry model of Scottish knitwear producers. *Admin. Sci. Quart.* 40(2):203–227.

Posen HE, Keil T, Kim S, Meissner FD (2018) Renewing research on problematic search—A review and research agenda. *Acad. Management Ann*. 12(1):208–251.

Rahmandad H (2008) Effect of delays on complexity of organizational learning. *Management Sci.* 54(7):1297–1312.

Ref O, Shapira Z (2017) Entering new markets: The effect of performance feedback near aspiration and well below and above it. *Strategic Management J.* 38(7):1416–1434.

Repenning NP, Sterman JD (2002) Capability traps and self-confirming attribution errors in the dynamics of process improvement. *Admin. Sci. Quart.* 47(2):265–295.

Shinkle GA (2012) Organizational aspirations, reference points, and goals: Building on the past and aiming for the future. *J. Management* 38(1):415–455.

Sinclair G, Klepper S, Cohen W (2000) What's experience got to do with it? Sources of cost reduction in a large specialty chemicals producer. *Management Sci.* 46(1):28–45.

Sitkin SB (1992) Learning through failure: The strategy of small losses. Cummings LL, Staw BM, eds. *Research in Organizational Behavior*, Vol. 14 (JAI Press, Greenwich, CT), 231–266.

Srinivasan R, Haunschild P, Grewal R (2007) Vicarious learning in new product introductions in the early years of a converging market. *Management Sci.* 53(1):16–28.

Starbuck WH (2009) Cognitive reactions to rare events: Perceptions, uncertainty, and learning. *Organ. Sci.* 20(5):925–937.

Strang D, Macy MW (2001) In search of excellence: Fads, success stories, and adaptive emulation. *Amer. J. Sociol.* 107(1):147–182.

Strang D, Patterson K (2014) Asymmetries in experiential and vicarious feedback: Lessons from the hiring and firing of baseball managers. *Sociol. Sci.* 1(May):178–198.

Strike VM, Rerup C (2016) Mediated sensemaking. *Acad. Management J.* 59(3):880–905.

Tamuz M (2001) Learning disabilities for regulators: The perils of organizational learning in the air transportation industry. *Admin. Soc.* 33(3):276–302.

Thaler RH (2011) Show us the data. (It's ours, after all.) *New York Times* (April 23), http://www.nytimes.com/2011/04/24/business/24view.html.

Thompson P (2007) How much did the liberty shipbuilders forget? *Management Sci.* 53(6):908–918.

Thompson P (2012) The relationship between unit cost and cumulative quantity and the evidence for organizational learning-by-doing. *J. Econom. Perspect.* 26(3):203–224.

Turner SF, Cardinal LB, Burton RM (2017) Research design for mixed methods: A triangulation-based framework and roadmap. *Organ. Res. Methods* 20(2):243–267.

U.S. Food and Drug Administration (2016) MAUDE: Manufacturer and User Facility Device Experience. Accessed September 13, 2016, https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/TextSearch.cfm.

Vaughan D (1996) *The Challenger Launch Decision: Risk Technology, Culture, and Deviance at NASA* (University of Chicago Press, Chicago).

Vissa B, Greve HR, Chen W-R (2010) Business group affiliation and firm search behavior in India: Responsiveness and focus of attention. *Organ. Sci.* 21(3):696–712.

Wooldridge JM (2000) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).

Zelner BA (2009) Using simulation to interpret results from logit, probit, and other nonlinear models. *Strategic Management J.* 30(12):1335–1348.

Zollo M (2009) Superstitious learning with rare strategic decisions: Theory and evidence from corporate acquisitions. *Organ. Sci.* 20(5):894–908.

David Maslach is an assistant professor of management at the College of Business, Florida State University. He received his Ph.D. from the Ivey School of Business at Western University. His research interests focus on how organizations learn, innovate, and respond to failure in advanced technologies.

Oana Branzei is the Donald F. Hunter Professor of International Business and an associate professor of strategy at the Ivey Business School at Western University. She received her Ph.D. from the University of British Columbia. Her research interests include the prosocial foundations, forms, and functions of business as agents of positive social change and the relational microprocesses of value creation, capture, and distribution.

Claus Rerup is a professor of management at the Frankfurt School of Finance and Management, Germany. He received his Ph.D. in organization theory from the Aarhus University School of Business and Social Sciences, Denmark. His research focuses on organizational sensemaking, routine dynamics, and organizational learning, with a particular interest in how these processes unfold across time and boundaries.

Mark J. Zbaracki is an associate professor of general management at the Ivey Business School at Western University. He received his Ph.D. from Stanford University. His research interests include the social construction of managerial practices, strategic decision making, strategy implementation, routine dynamics, and organizational learning.