The Cycle of Doubt:

How Scientists Cope with Uncertainties in Producing Data

Seokkyun Woo[1]

John P. Walsh[2]

December 2023

[1]Graduate School of Science and Technology Policy, KAIST

[2]School of Public Policy, Georgia Institute of Technology

Abstract

How do scientists address failures when producing experimental data? While the social dynamics of failures at the epistemological level – such as questioning or defending theories upon failures – are well examined, failures "at the bench" in day-to-day laboratory work have been less understood. Moreover, failures from the bench often precede the aforementioned epistemological interplays involving evaluation and readjustment of theories and hypotheses. We explore these "low-level data generation" processes based on ethnographic observations from material science labs to understand the work of scientists addressing data production. Analyzing these failures provides a strategic site for the sociology of science as it not only addresses potential bottlenecks in production of science but also reveals problem-solving processes, skills, and the interplay between epistemological and material dimensions and the degree to which the knowledge is situated. We find that scientists' *situated knowledge* of empirical data production work is essential in identifying and searching for causes and solutions to failures. Secondly, due to the situated nature of data production, we find that necessary information for addressing these failures is primarily communicated and shared among lab members who are engaged in similar tasks. Our last main finding reveals an interesting aspect of scientists' search behavior. We find that scientists tend to internalize failures initially, blaming themselves, before attributing these failures to external factors. Based on our in-depth observations and conceptualizations, we present a model named the "Cycle of Doubt," which generalizes the failure-coping process. We provide a detailed discussion of the theoretical and practical implications of our findings.

Keywords: Data Production Failure, Situated Knowledge, Lab Ethnography, Attribution Behavior

The Cycle of Doubt: How Scientists Cope with Uncertainties in Producing Data

1. Introduction

The production of data to study nature is a critical aspect of modern science. However, data production work, such as preparing samples, conducting chemical syntheses, and operating instruments, is riddled with potential failures (Delamont and Atkinson 2001; Lynch 1985; Shapin and Schaffer 1985; Wylie 2019). These can emerge in a variety of, often unexpected, ways, creating a burden for scientists as they must make a series of decisions to try to overcome production interruptions and failures (Barley and Bechky 1994; Doing 2004). Studying failure in science has attracted many scholars as it represents a critical juncture where theories can be readjusted and reexamined. For example, prior studies have focused on the epistemological aspects of failure, exploring how scientists reassess and modify hypotheses and theories upon experimental failures (Kuhn 1962; Latour and Woolgar 1979). While understanding the epistemological context of failure is crucial to understand how scientists reconfigure their interpretation of nature, it is important to note that these intellectual deliberations are often preceded by actual data production failures occurring in the laboratory setting, suggesting the necessity to directly observe scientists addressing data production failures in their day-to-day work setting.

Recognizing that data production failure can pose a significant bottleneck in the production of science, and furthermore, that studying these failures can offer important insights into understanding knowledge production and social dynamics in scientific endeavors, this paper provides detailed accounts of how scientists address data production failure - the failures and interruptions while producing data for their research. We are especially interested in how data production failure is constructed, recognized, and addressed and the extent to which this process can be shaped by scientists' current and past interactions with their social and physical settings. In doing so, we aim to identify the kinds of skills employed, understand the role of the lab structure, and examine how these failure-coping processes might be affected by the attribution of failure and the specific characteristics of failure-generating tasks, and the role of situated knowledge in that attribution. Furthermore, we ask how scientists search for possible causes and solutions when encountering failures. Based on our in-depth observations and conceptualizations, we present a model named the "Cycle of Doubt," which generalizes the failure-coping process.

Addressing these questions requires us to directly observe the production settings in which scientists generate data. Thus, using a sociology of work perspective (Barley and Bechky 1994; Barley and Kunda 2001; Becker 2002; Orr 1996), we conducted two years of ethnographic observation of two material science labs at a research-intensive university in the US. Direct observations of failures and the failure-coping process in real time as well as informants' accounts of recent and on-going failure experiences allowed us to provide a detailed examination of how laboratory scientists handle data production failures in their day-to-day work. Our findings first suggest that data production failures are very common and occur for a variety of reasons, such as instability of a production process, instrument malfunction, material quality issues, or inadequate procedure. The failure-coping process involves the identification of data production failure, making sense of failure, searching for potential causes and solutions, and iterations of this process until scientists "solve" the problem or simply move on to the next tasks. This process, which we describe as the "Cycle of Doubt," is articulated in detail in the Findings section. What we found intriguing was the high frequency of failures and the substantial time spent addressing them. Moreover, identifying the exact cause of the failure was often difficult, and in some cases not even a primary goal of the problem solving process. In the face of the highly uncertain and unstable production process, useful feedback during problem-solving failure was scarce and ambiguous. In such an environment, we find three intriguing features of the failure-coping process.

Firstly, we find that scientists' *situated knowledge* of empirical data production work is essential in identifying and searching for causes and solutions to failures. By "situated," we mean that scientists' working knowledge to produce data is situated around their production setting (Lave and Wenger 1991; Tyre and Von Hippel 1997). Situated knowledge includes the ability to precisely make use of local information to detect, make sense of, and search for solutions to data production failures. When the information generated around failure was not sufficient, some scientists would devise sub-experiments to generate non-research-related information, a key feature of the failure-coping process. However, it is important to note that the ability to make use of situated knowledge, in particular, devising sub-experiments, varied by production settings. From our observations in labs staffed by graduate students and postdocs, this situated knowledge was often used with formal scientific knowledge to address data production failures. We thus anticipate that labs with varying numbers of technicians, whose main task is to maintain instruments and produce experimental data (Barley and Bechky 1994; Latour and Woolgar 1979), will display different compositions of formal and situated knowledge among their lab members.

Secondly, because of the unstable production environment and the importance of situated knowledge, we find that scientists who share data production with other members of the lab are more effective in addressing data production failures. Due to the stickiness of the information (Von Hippel 1994), communications and knowledge spillovers were much more common among scientists who work with the same materials and instruments in a shared space. This finding aligns with prior findings that emphasize the importance of dense networks with strong ties for transmitting tacit knowledge (Hansen 1999), providing important insights into the organization of work in science.

Lastly, we find an intriguing behavioral pattern that governs scientists searching for potential attributions to failures. Our participants tended to initially search for failure attributions over which they had complete autonomy, which we categorized as an *internal attribution*. Examples include questioning their direct roles in data production failure. Only a handful of failure-solving processes started with *external attribution*, such as questioning published techniques or the quality of materials prepared by collaborators. While scientists internalizing failures may be attributed to enculturation (Delamont and Atkinson 2001; Wylie 2019), social norms (Merton 1973), and their prior beliefs (Dunbar 1995; Klahr 2002), it is important to note that such internalization of failure can often explain the delay in identifying false research that leads to a replication and reliability crisis (Reich 2009).

This paper first reviews relevant studies from STS, sociology of work, and cognitive science that provide useful insights for our research questions and methodologies. We then discuss our methods and data collection, followed by our findings on how scientists address data production failures. We also introduce a behavioral model developed based on our observations, which we call the "Cycle of Doubt." We then discuss the implications of our findings, including for the organization of work in science and the potential roles of AI in improving data production.

2. Literature and Research Questions

2.1 Production of data as craft and routine

Data production, which we define as *tasks involving sample preparation, making materials, and running instruments by bench scientists,* is foundational in empirical science. In an analysis of contribution statements for almost 80,000 authors from PLOS ONE, Sauermann and Haussler (2017) find that half of the authors, and 86 percent of first authors, were engaged in data production. Bench scientists commonly face frequent and unpredictable production failures and interruptions in producing this data, exacerbating the time demands (Cambrosio and Keating 1988; Delamont and Atkinson 2001; Fujimura 1987; Jordan and Lynch 1992; Lynch 1985; Peterson 2015; Wylie 2019). Data production done in an "open system," such as chemical synthesis, is notoriously troublesome. Even in what may seem a well-controlled environment, scientists often must "babysit" instruments by monitoring potential signs of failures. The following quotes from these prior studies provide some examples of the time involved and types of data production failures. *"In the beginning, the first task at hand was actually to express the protein and that was where we hit the roadblock. As a consequence, it took her two-and-a-half years to solve a problem that was originally envisaged as a starting point for the doctorate"* (Delamont and Atkinson 2001). *"Now I've gone through a lot of the problems. I've had it go wrong, I've lost samples, it's cost me a whole week, I've had to start all over again"* (Jordan and Lynch 1992). Thus, it is not surprising that a substantial share of graduate training in laboratory science involves getting familiar with the procedures and materials at the bench (Campbell 2003; Delamont and Atkinson 2001; Peterson 2015). These studies have emphasized the importance of having (Peterson 2015) or acquiring (Delamont and Atkinson 2001; Doing 2004) craft and tacit knowledge (Collins 2010). Even tacit knowledge may not be sufficient when the cause of failure is related to the nature of the unstable data production process (Lynch 1985). For example, producing a good crystal from a complex molecule in structural biology can take months or even years (Ramakrishnan 2018). Thus, examining how scientists address technical troubles allows us to understand different problem-solving skills and how contextual factors shape this process.

It is thus natural to ask to what extent can the production of data in laboratory science be rationalized. By rationalization of data production, we mean formalization and standardization in the work of producing data, such as the codification of experimental protocols or developing/purchasing research instruments that would increase precision by partly automating otherwise precarious tasks. One way to rationalize work is to embed such tasks into organizational memory in the form of organizational routines (Argote and Darr 2000; Cyert and March 1963; Nelson and Winter 1981; Simon 1947). A somewhat related concept from the STS perspective is the "standardized package" (Fujimura 1987; Fujimura 1988; Fujimura 1992), which provides standardized practices of techniques, research materials, and instruments. Another related concept is the "black-box" (Latour 1987; Latour and Woolgar 1979), which provides theory and consensus laden practice in the forms of standardized research instruments or techniques. To the extent that these work practices become a part of the organizational routine, one can expect significant reductions in training costs and time incurred by research labs that are often plagued by the depreciation of organizational knowledge due to high turnover rates among members (Knorr-Cetina 2009).

However, rationalizing data production work can be difficult for several reasons. One major obstacle is physical constraints (Barley and Bechky 1994; Bruyninckx 2020; Cambrosio and Keating 1988; Jordan and Lynch 1992). For example, chemical synthesis in an "open system" is prone to interference from local parameters, such as atmospheric gas, room temperature and humidity, type of tools they use, conditions of materials, and even seemingly trivial permutation of tasks. Successful data production in such a setting means having lower variance over repeated trials. Rationalization of work may not necessarily be more effective in "closed systems", such as those that are partly automated through pre-programmed chemical reactors and instruments, as previous studies on the work of technicians and engineers suggest that production failures are common, and codified work procedures provide little value when addressing technical failures (Barley 1988; Kusterer 1978; Orr 1996). This could be because many standardized practices are often designed for specific uses and may not work in many situations and settings different from their original purpose (Lave and Wenger 1991; Suchman 2006; von Hippel and Tyre 1995). In the case of scientists, since they are pushing the boundaries of knowledge, they often find themselves pushing against the limits of the instruments. Also, the normative emphasis on originality (Merton 1957) may lead scientists to work on non-redundant research, which is often accompanied by local variations and improvisations of techniques (Jordan and Lynch 1992; Knorr-Cetina 2009). Moreover, while automation can replace some manual tasks, scientists need to constantly address the machines' erratic behaviors (Barley and Bechky 1994; Bruyninckx 2020; Ribeiro et al. 2023). Sometimes, scientists may view rationalizing data production tasks as trivial and costly. Even if they have successfully standardized data production tasks, they may not find publishing their methods worthy of their time and effort (Cambrosio and Keating 1988). Thus, many important data production techniques remain local and tacit, facing a risk of disappearing with a turnover (Argote and Darr 2000; Knorr-Cetina 2009).

2.2 Situated knowledge in failure-coping process

Unlike industrial manufacturing work, laboratory science often involves combining materials and techniques in unusual ways (Barley and Bechky 1994), making standardized techniques less useful for labs that are conducting novel research (perhaps especially novel materials). Situated learning theory argues that knowledge used to solve a problem is situated in the sense that it depends on the social and physical context in which the problem occurs (Lave and Wenger 1991; Suchman 2006). This theory provides useful insights into the management of troubles in such work settings (Barley and Bechky 1994; Barley 1988; Lave and Wenger 1991; Orr 1996; Suchman 2006; Tyre and Von Hippel 1997). The sociology of work literature (Harper 1987; Kusterer 1978) describes the role of situated knowledge ("working knowledge") in managing technical troubles in a production site. For example, Kusterer (1978)'s study of manufacturing workers on a production line showed that both material-specific and machine-specific knowledge is necessary for uninterrupted production. The acquisition and application of this working knowledge are further elaborated by prior studies on the work of technicians (Barley and Bechky 1994; Barley 1988; Orr 1996). Just as technicians are responsible for transforming physical objects into symbolic representations (Whalley and Barley 1997; Zuboff 1988), laboratory scientists transform organisms and chemical materials into a symbolic representation as data via various processes and scientific apparatus (Barley and Bechky 1994; Latour and Woolgar 1979). Managing these tasks requires careful attention to the idiosyncrasies of materials and instruments. What is important is that this working knowledge is used to make sense of produced data, in particular, for maintenance and diagnoses of troubles (Barley and Bechky 1994; Whalley and Barley 1997). Thus, scientists' knowledge generated by performing such tasks, whether they are in the form of tacit or codified knowledge, is situated in the physical context in which productions occur.

2.3 Situated knowledge in the production of science

Situated knowledge is particularly important in laboratory science due to two different forces that create variations in techniques and methods of producing data. These variations can arise from the normative emphasis on originality (Merton 1957), which discourages scientists from pursuing repetitive projects, resulting in increasing variations in the data production settings both within and across laboratories. Variations can also happen during the adoption of standardized methods as lab members transform standardized production methods into situated knowledge in the forms of local and informal routines (Cambrosio and Keating 1988; Jordan and Lynch 1992). These frequent changes in the production setting include constant modification and improvisation of scientific apparatuses and the production and testing of new material. Under such temporal instability, situated knowledge may become essential for maintaining uninterpreted data production. A key insight from situated action theory is that how data production failure is constructed, recognized, and addressed may not be independent of scientists' current and past interactions with their social and physical settings.

Previous studies on lab technicians (Barley and Bechky 1994; Bruyninckx 2020; Doing 2004) document how situated knowledge is put into action to solve technical troubles in bench science. For example, lab technicians may employ a wide range of available heuristics, which are embedded in and acquired through interactions with their production settings, to detect data production failures. It is interesting to note that these studies have primarily focused on technicians working in labs with a division of labor that clearly distinguishes technicians' work from that of scientists. Meanwhile, many basic science labs in US research universities are staffed with postdocs and graduate students who may be responsible for both technical and scientific aspects of the laboratory work while solely being evaluated by their formal scientific knowledge (Hackett 1990; Hagstrom 1964). It is unclear how situated knowledge addresses data production failures in labs without dedicated technicians. In this sense, our observation of department-level material science labs may provide an interesting opportunity to observe how situated knowledge and formal scientific knowledge (or both) are used to address data production failures.

2.4 Structure of the failure-solving process

Having discussed the role of situated knowledge, we now turn to the structure of a failure-solving process. This refers to the different stages involved in addressing data production failure, including failure detection, attribution, solutions, and learning from failure. Indeed a previous study (Barley and Bechky 1994) suggests that the failure-solving process goes through iterations of representation updates and reframing of the problem. However, there are still important questions related to understanding the structure of failure-solving processes and how the process may be influenced by the attribution of failure and the specific characteristics of failure-generating tasks. It is also important to understand the conditions that produce a faster transition from a wasteful search to a more productive search. To better understand this issue, we discuss some useful insights and findings from the cognitive science literature on problem-solving in the context of scientific discovery (Klahr 2002; Klahr and Simon 1999; Simon, Langley and Bradshaw 1981).

These studies have shown that solving scientific problems involves selective search using either domain-specific or general heuristics over a large space of possibilities, suggesting that the process of scientific discovery is not entirely unique but instead is a special case of the human problem-solving process (Simon 1992). Dunbar (1995)'s study shows that scientists use different heuristics to address inconsistent evidence or experimental failures, with substantial heterogeneity in the uses of different search modes by type of problem and social structure of labs. The study points out that scientists tend to search locally when addressing experimental problems (what we call data production failures). Dunbar also reports that around 60% of the experiments had experienced technical problems, highlighting the need to address data production failure to increase the pace of science. Studies by Gorman (1986) and Klahr and Dunbar (1988) found that the presence of data error can slow down problem-solving by introducing another layer of uncertainty, leading to potentially wasteful replications of flawed experiments (a problem solving heuristic of "repeat and hope for the best"). Because of high uncertainty, it is plausible that some unknown noise caused the failure and repeating the same process might, this time, produce a successful outcome. It is unclear when scientists should break out of this repeat-and-hope-for-the-best mode and change search strategies to overcome data production failure.

The prior studies from the STS, sociology of work, and cognitive science literatures suggest that data production failure is common yet difficult to solve, and that situated knowledge may be crucial in solving failures. Moreover, understanding the structure of the failure-solving process may be key to improving the data production process, which is essential for increasing the pace of science. Building on this perspective, we ask the following specific questions about scientists addressing data production failures: *1) how do scientists detect data production failures? 2) what are the different types of data production failures? 3) what are the common causes of the failures and how do scientists construct failure attributions? 4) what are common solutions, and how are they applied? 5) what are the conditions that generate faster solutions?* The next section provides a detailed description of our empirical methods.

3. Methods and data

3.1 Methods

We conducted two-years of ethnographic observation and unstructured interviews with doctoral students, post-docs, and faculties in materials science labs. Ethnographic studies are a long-established method of understanding the social activities of scientists (Delamont and Atkinson 2001; Fujimura 1987; Knorr-Cetina 1981; Latour and Woolgar 1979; Owen-Smith 2001). In contrast to previous science studies that focused on social interaction in lab meetings, our observations took place in the laboratories where scientists' work of producing data normally occurs (Barley and Bechky 1994; Latour and Woolgar 1979; Lynch 1985). We use this approach for the following two main reasons. First, while observing social interactions may be important to understand the social construction of knowledge, data production generally precedes any meaningful discussions around it. Thus, it is important for us to directly observe the work of scientists to gain a contextual understanding of their work processes (Orr 1998) in order to avoid "a version of the event which has been eroded of all contingent circumstances" (Latour and Woolgar 1979). The second benefit of our approach comes from the prevalence of data production failures in laboratory science, which spares us from picking a project and observing it from start to end. While we did not observe the entire life cycles of all projects, the external validity of research is enhanced by visiting two different labs, spanning multiple projects. Since these labs varied in size, structure, and studied materials, this also allows us to observe potential organizational heterogeneities in searching for solutions (Dunbar 1995).

3.2 Data

Our data come from two rounds of ethnographic observations in two material science labs from a research-intensive university in the US. The first round of observations lasted from November 2018 to February 2020, cut short due to the Covid Pandemic. The second round of observations lasted from February 2021 to March 2022. The observations focused on failure-solving activities while running experiments. Observational data collection relied primarily on the "thinking-aloud protocol" (Newell 1967) where scientists were asked to talk out loud describing their failure-solving process. While our field visits lasted, on average, around two hours, we followed four scientists from two different labs for a considerable period of time. Thus, we were able to come back to any data production failures that weren't resolved (many failures took days, weeks, and even months) at the site by following our participants for months and years. Informal interviews were periodically conducted at the site asking them how they addressed the failures. These observations were recorded through notetaking and tape-recording and transcribing. These field notes were used to generate failure-solving scripts that were coded and analyzed to understand how scientists interpret and respond to data production failures. The failure-solving scripts were coded in an Excel spreadsheet. The spreadsheet includes the entire entry of data production failures that we observed at the site or reported by our participants. For each data production failure entry, the following items were coded: date, lab, location, project, participant name, raw texts from field notes, instruments and materials involved in the failure, whether the failure was observed on-the-spot versus reported later by our participants, the familiarity of the failure-generating tasks, failure attribution, and the solution used to solve the failure. Further analysis that required examining the occurrence and co-occurrence among these items was done using the author's own Python script.

3.3 Sites

We chose materials science labs as the research site for several reasons. Not only is materials science unique in that the research program involves studying and creating new materials, unlike other disciplines that focus on understanding natural phenomena, but the high level of difficulties and uncertainties involved in creating and testing materials also suggests that materials scientists are situated in a problem-solving space that must accommodate both the logics of scientific discovery and the logics of troubleshooting technicians. This dual nature, coupled with the fact that there is high visibility of the work, makes the material science labs a strategic site to study data production failures.

The two material science labs that we visited provide interesting variations in research areas, routine work, and how they organized their work. In the material science department, labs are broadly classified by their scientists into either "wet" labs or "dry" labs, depending on whether the material system they study has more fluid (wet) or solid (dry) characteristics. The first lab is involved in research related to producing and studying synthetic and biological polymers and particles in fluids. By this standard, our participant called their lab a typical "wet" lab, with much of their work involving complex chemical synthesis involving fluid materials, and hence we will refer to this as Wet Lab. During the period of our field visit, Wet Lab had 6 members: 1 faculty supervisor, 1 postdoc, 3 doctoral students, and 1 master's student. The material scientists of Wet Lab spend most of their bench work in three different physical spaces: the preparation room, synthesis room, and optics room. The preparation of samples and materials that do not involve chemical synthesis was done in the preparation room. Chemical synthesis was done in the synthesis room where each member had their own designated fume hood. As with most of the other material science labs, carefully made materials would undergo various characterization tests. One specialty of Wet Lab was their ability in making new characterization techniques using optical instruments, which were done in the optics room. The optics room also housed various lab-built characterization instruments. Other standard off-the-shelf characterization instruments were located in the preparation room.

One interesting characteristic of Wet Lab was that its member pursued their own projects, which were quite distinct and had low task overlap with their other lab members. For example, while one participant in Wet Lab would spend most of her time performing chemical synthesis, another participant would spend most of his time devising new characterization techniques. However, this was not indicative of the lab's division of labor, instead, this was more indicative of their low task overlap. In fact, one participant from Wet Lab told us that his supervisor demands every member to know a little bit of everything, especially, tasks involving chemical synthesis, building and re-building research instruments, characterization of molecules, and programming problems.

An equal amount of field observation came from the second lab. By material scientists' standards, this second lab was a "dry" lab whose research involves making and studying organic-inorganic hybrid materials, and hence we will we refer to this as Dry Lab. Dry Lab was a mid-size lab with 10 members: 1 faculty supervisor, 2 postdocs, and 7 doctoral students. The lab also had a constant influx of undergraduate research assistants. Unlike Wet Lab, Dry Lab's routine work was highly concentrated on chemical reactions using large lab-built reactors. The lab was primarily interested in developing organic-inorganic hybrid materials using vapor phase infiltration (VPI), a variant of atomic layer decomposition (ALD). Most of the benchwork from Dry Lab was done in three main physical spaces: the reactor room, characterization room, and preparation/synthesis room. Most lab members spend their time in the reactor room as it houses various chemical reactors built or modified by lab members. While there were many variations in their design, all reactors had a chamber, a large metal container into which a base material (substrate) was put. For example, in a VPI reactor, organic material is placed into the chamber, and inorganic material (precursor) is injected into the chamber through a pneumatic system based on a pre-programmed method. The variation in design among reactors results from the type of substrates that researchers want to be "baked", and more importantly, the type of data generated during the reaction process. Members of Dry Lab also spent a substantial amount of their time in the characterization room. Unlike their reactors, most characterization instruments were "off the shelf" types, purchased from instrument manufacturers. This reflects Dry Lab's research orientation in making new materials, which contrasts with Wet Lab's research orientation which encompasses the devising of new characterization methods. Unlike Wet Lab, the members of Dry Lab spent little time in the preparation/synthesis room. While there were a few fume hoods in the room, lab members would rarely perform any chemical syntheses, as most of base materials (substrates) they used were made by their collaborators from the Chemistry department or other "wet" labs from the Material Science department.

In terms of organization of the work, Dry Lab was highly specialized as a whole as the lab was mainly focused on making novel functioning materials using specialized chemical reaction methods. The work overlap was high as lab members would often share one of their core instruments. Meanwhile, specialization was observed in terms of how each lab member was trained in operating different characterization techniques. Because Dry Lab would often need to use shared facilities to have access to high-functioning characterization equipment, such as X-ray photoelectron spectroscopy (XPS), selected members were trained to operate them. Such specialized training reflects a selective division of labor operated in Dry Lab. Where lab members would jointly share chemical reactors (making of materials), some characterizations (understanding of materials) were performed only by specially trained members. The relatively high overlap of tasks while retaining some division of labor contrasts sharply with Wet Lab's organization of work, characterized by low work overlap and the lab's emphasis on every member knowing a little bit of everything. Both labs are interested in making and understanding novel materials, but they differ in their research topic, work routine, and work organization. The routine work done in Wet Lab can be considered as an "open system", where the data production process is highly vulnerable to external interferences (i.e., chemical synthesis done in a fume hood). Meanwhile, much of the data production from Dry Lab is conducted within chemical reactors, or a "closed system."

It is important to emphasize that all members across both Dry Lab and Wet Lab were highly skilled and professional in their roles as experimental scientists. The "failures" observed during our study were not a result of any incompetence or lack of diligence on their part. In fact, there was a state of consensus among most prominent material scientists that such challenges are an integral and often expected component of scientific works. As one piece of corroborating evidence, we had the opportunity to present our initial findings to a National Academies study panel on data analytics for material science. The panelists agreed that data production failures were very common bottlenecks in their own labs. While there are likely individual differences in problem solving skills, the routineness with which people in these labs, and in the field more generally, approached these failures suggested that it is reasonable to focus on the contextual and structural factors, such as task characteristics, task familiarities, and the degree of task overlap, for understanding the problem solving process (Dunbar 1995; Newell and Simon 1972).

4. Results

4.1 Construction of failure

Recognizing failures

Our observations suggest data production failures were common. We observed 38 instances of data production failure during our 36 on-site observations. This number excludes those that were brought up during discussions with our participants. Since our field visits lasted around 2 hours on average, our participants were interrupted once every 2 hours during their benchwork. Thus, troubleshooting failures took up a significant portion of their time. Some failures were easily resolved, while others took days or months to be resolved.

Data production failure could happen due to a wide range of reasons, including but not exclusively due to sample impurity or contamination, instability of the production process, instrument failure, and inadequate experimental design. As these failure attributions were often identified in retrospect, troubleshooting often started with recognizing that their data production process may have failed. Detecting data production failures involved both situated and formal scientific knowledge. For example, our participants would recognize data production failures by making use of their sensory cues (Barley and Bechky 1994; Bruyninckx 2020) or by noticing anomalies in figures, graphs, and data tables. The ability to recognize data production failures varied according to our participants' task experience with specific instruments or material systems (Kusterer 1978; Simon 1947; Suchman 2006). Some failures were relatively easy to detect, such as when instruments signaled warning signs or simply stopped running. Materials could burn, break, or turn into unexpected colors. Characterization results could also drastically deviate from the expected range of values, shapes, and patterns (for example, seeing a flat line across all values of the x-axis). Meanwhile, other failure recognitions would require a more nuanced situated understanding of the production process. For example, experienced participants performing unstable chemical synthesis were able to utilize a wide range of visible cues embedded in their production settings, and their higher order relationships to monitor potential data production failures.

Making sense of failures

Previous cognitive science studies provide useful concepts to describe the perception of failures. Klahr and Dunbar (1988) proposed a dual search process for scientific discovery, where hypotheses are generated in the hypothesis space and experiments are selected from the experiment space. The study reveals that the prior knowledge and the surrounding empirical context impose strong biases in generating and evaluating hypotheses. Inspired by Newell and Simon (1972)'s definition of the structure of problem space, we classified failure-generating tasks as either familiar or unfamiliar based on how clearly our participants understood the production recipes, procedures, and goals.

Failures from familiar tasks

While our participants focused on creating and studying novel materials, some processes of making, characterizing, and testing these materials were familiar to participants as they followed some regular task sequences. For example, Dry Lab focused on understanding the fundamental thermodynamic properties governing various chemical vapor deposition processes, involving routine tasks such as spin coating substrates, infiltrating metal oxides into substrates using lab-built instruments, and obtaining inscriptions from materials with various characterization techniques. They also had experience building their own reactors, which involved technical knowledge unrelated to formal scientific training. These repetitive tasks and instrument building experience served as a training ground for lab members to improve their craft and develop heuristics for detecting and attributing failures. Similar repetitive tasks were observed in Wet Lab, such as routinely creating stocks of synthesized macromolecules used in different research projects. Lab member's situated knowledge accumulated with formal scientific training, including advanced courses on characterizations such as spectrometry and microscopy. Thus, a substantial share of their works was familiar to them in the sense that they had developed both situated and formal scientific understanding of the materials and production process.

Our analysis from coded field notes shows that around 55% of the instances of failures happened while performing familiar tasks, indicating that familiarity with tasks does not necessarily prevent data production failures. Familiarity, however, led to lower variance and the failures from familiar tasks did not surprise the participants as some of these tasks were known to be notoriously tricky. We provide one example from a senior doctoral student in Web Lab who spent many years synthesizing macromolecules. Her research involved understanding the various properties of polypeptides, a material system that she often produced in large quantities. The following steps summarize one of the intermediate steps of polypeptide synthesis: extraction of the organic phase of silica.

"She retrives a flask from the oven for her moisture-sensitive experiment. She adds water into the silica solution, which creates a layer that separates the organic phase (showing yellowish color) and the aquatic phase (clear solution). She gently taps her flask and says this will help more of the organic phase to rise to the top. She pours the sodium bicarbonate solution into the flask. She opens the valve, and the aquatic phase (that was sitting at the bottom) is poured into a beaker. She then adds sodium bicarbonate solution again into her flask. She shakes it, and it turns into a lemonade color. After 5 minutes, the layer has formed again. She quickly inserts her solutions into the plastic vacuum bag, which had been purged with inert gas. After adding drying agents into her solution, she then pours it into another flask down through a paper filter. An air pump is connected to the other flask to accelerate the filtering process. She says all this process must be done quickly to minimize exposure to oxygen."

As seen from this field note, the production of polypeptide synthesis is highly sensitive to both time and the surrounding environment. One critical step in the synthesis must be performed in a vacuum to minimize oxygen contact, which she tried to minimize by installing an air pump to accelerate the filtering process. A few days later, characterization data revealed that the polypeptide she made had low yields. When questioned about the cause of the failure, she attributed it to the precarious production process, which was particularly exacerbated by time and air sensitivity.

"Depending on the reaction, you can get high molecular weight and low molecular weight ... Because it depends on the reaction, the reaction for making polypeptides is moisture and air-sensitive. So, if you don't wash (purify) very well, you can get a really low molecular weight yield. So, it depends on how pure your solution is."

Data production failures were also commonly observed from "routine" tasks in Dry Lab. Many members of the lab used lab-built chemical reactors that could perform various chemical vapor deposition reactions. The entire process could be automated with user-friendly software, allowing them to replicate any previously run chemical reaction. While automation allowed them to focus on other tasks, they would regularly stop by the reactor room to monitor temperature, pressure, and power gauges. This was because reactors could frequently fail in different ways and monitoring was necessary to identify the cause of failure and avoid wasting valuable time, as the entire reaction could take 2-7 days.

Failures from unfamiliar tasks

The remaining 45% of the data production failures occurred in unfamiliar tasks where production recipes, procedures, and goals were not clearly defined and understood by our participants. In some cases, production tasks had clearly defined production recipes and goals, but the procedures were ambiguous. For example, one participant from Wet Lab was trying to replicate a chemical synthesis of cobalt-silica from published research. She said that producing a spherical silica particle coating around cobalt turned out to be a tricky process, although the production recipes and the procedure were codified in a published manuscript. As seen from the field notes below, our participant and her colleague struggled with the replication process due to many unknown experimental parameters that were omitted from the paper.

"So, I am basically trying to replicate this process (synthesis of silica-cobalt)… However, they (paper) do not explain all the processes, which makes it hard to replicate… The success rate is like 1-2 out of 10 … This is not for the entire process but for the final stage … Synthesis is highly influenced by your environment. It could be influenced by the surrounding humidity. So, for example, it could work in Boston but not in Atlanta or San Francisco. Also, some of the "little" things are not written down. Those would be like second-degree problems for them. For example, they did not write down how much time they degassed when adding cobalt into water. Maybe for them, this wasn't a really important part. One of our grad students contacted a lab member of this corresponding author to get more information."

Producing cobalt-silica was unfamiliar for her not only because of her inexperience in producing it, but the limited information she had from the published paper, which was later recovered by contacting the original authors of the paper. Her experience illustrates a well-documented issue with the incomplete codification of lab-specific knowledge (Cambrosio and Keating 1988; Delamont and Atkinson 2001; Star 1983), which may be an important source of replication problems. But it also shows that the dichotomy between codified and tacit knowledge may be incomplete to explain her struggle. From Wet Lab's point of view, what was missing from the manuscript included a part of the original authors' situated knowledge, more specifically, parameters from the physical site on which the original synthesis was conducted. Thus, the problem of omitted experimental parameters was rather a matter of choice made by the original authors about what was important to report, rather than whether such knowledge can be codified.

In other cases, goals may be clear but finding the right recipes could make the task difficult. This is well illustrated by a senior doctoral student from Wet Lab who was devising a novel method of characterizing polymers by using an off-the-shelf instrument to detect conformational changes. While he had already demonstrated the method's usefulness with polypeptides, he was now searching for other material systems to show the method's general applicability. He was now in unfamiliar territory.

Material scientists encounter both familiar tasks as well as unfamiliar tasks in their work. Familiar tasks involve established methods such as known chemical synthesis to produce commonly used substrates, operating specialized chemical reactors, or producing data with established characterization instruments. Unfamiliar tasks often involved a deviation from familiar production recipes or procedures, in which the goal was to produce new findings. Data production failures were more or less equally observed from both familiar and unfamiliar tasks (although we do not have data on the proportion of effort devoted to each type of task).

4.2 Attribution space

Once data production failures were detected, our participants generated a number of potential failure attributions. Based on our observations, we classify potential causes of failures into six different categories: *instability, material quality, instrument, procedure, nature,* and *don't care*.

We then refer to the selected failure attributions by our participants as the attribution space. The construction of the attribution space is shaped by both our participants' underlying prior experience as well as the situated contexts surrounding the data production failures. Figure 1 illustrates the relative frequency of failure, categorized by failure attributions and task familiarity. We discuss each of these failure attributions in the order they are shown in Figure 1.

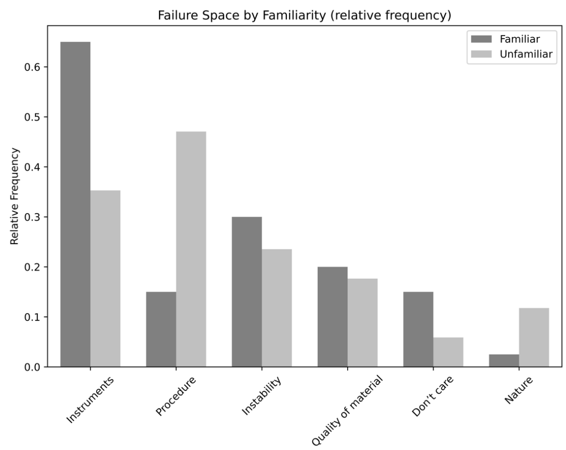


Figure 1. Failure space distribution by familiarity. Data comes from 82 instances of data production failures that were directly observed and retroactively recorded. Note that the relative frequency for each failure attribution is calculated by dividing the number of occurrences of a specific failure attribution by the total number of failure instances that sought either internal attributions (represented by dark grey) or external attributions (represented by light grey).

Instrument

Data production failures could arise from malfunctioning or incomplete understanding of instruments. Such issues are labeled as *instrument* failure. This can occur when instruments are pushed beyond their capacity or when improvising instruments by integrating multiple subcomponents purchased from different vendors. Interestingly, *instrument* failure was by far the most common failure attribution sought from our participants among familiar failure-generated tasks (see Figure 1), suggesting a tradeoff between using customized instruments for high-quality data and the greater risk of data production failures.

Procedure

Our participants often attribute failures to the inadequate experimental setup and try different variants of the existing procedures or change the material system. This type of failure attribution is categorized as *procedure* failure. *Procedure* failure shares some similarities with *instability*, which addresses inherent instabilities in data production. The key difference is that *procedure* failure can be attributed to the configuration of the data production setup that could be deliberately improved. Procedure attributions occur more often among non-standardized methods and reflect a continual improvement of the data production process in an unfamiliar space. The fact that procedure failures are much more common among unfamiliar tasks, as shown in Figure 1, corroborates this interpretation.

Instability

Much of data production in laboratory science involves inherently difficult production processes (Barley and Bechky 1994; Cambrosio and Keating 1988; Delamont and Atkinson 2001; Knorr-Cetina 2009). We categorize this type of failure as *instability* failure. Examples include highly unstable chemical synthesis done in an open system or an infiltration process involving complex chemical interaction, which often renders troubleshooting difficult as they cannot always determine whether the failures are due to their own actions or unobserved experimental parameters. Production processes that require precise movements, or lack of movement (steady hands), would also be characterized as instability (Owen-Smith 2001; Peterson 2015). The variations in the output of a golf swing might be the canonical case. Thus, our participants often considered *instability* a default attribution when encountering failures from unstable data production processes.

Material quality

Data production failures were also often attributed to materials or samples that were impure or contaminated. We categorize this type of failure as *material quality* failure. Wet Lab, which studied fluid materials, faced more contamination problems. Protocols such as using gloves, lab cleaning products, and rinsing glassware to remove residuals were often followed in Wet Lab. In addition, *material quality* failure could be detected during or after experiments using available heuristics such as unusual color, smell, or texture (Barley and Bechky 1994; Bruyninckx 2020) or catching unusual characterization outputs (Lynch 1985). These abilities to make use of cues embedded in their production setting were accumulated through their previous encounters with similar failures and their recollection of how local materials were prepared and handled. Interestingly, special attention was often given to purchased materials as their impurities could cause problems such as replication failure.

Nature

Sometimes, our participants attributed data production failures to their lack of understanding of the material or the production processes, which we categorize as *nature*. For example, having repeatedly modified the *procedure*, and checked the *material quality* and *instrument,* one participant from Wet Lab attributed the failure to his lack of full understanding of a complex interaction between material systems and his instrument. We find this *nature* category interesting because making this attribution and addressing it could reveal unexpected findings, such as research fraud, as in the case of the infamous Schön scandal (Reich 2009), or serendipitous discoveries, such as the discovery of the Pulsar (Burnell 2004).

Don't care

In some cases, when our participants could not come up with failure attributions, instead of attributing the failure to *nature (*meaning that there was an underlying but as yet not well understood natural phenomenon causing the failure*)*, they would attribute the failure to unknown technical difficulties, where understanding the difficulty would not significantly expand their understanding of science. We categorize this type of failure as *don't care*. Our participants often ignore this type of failure (meaning they did not try to find the cause) or apply heuristics to solve them, seeing little value in understanding their exact cause. *Don't care* attribution was more often observed from failure-generated tasks that were familiar to our participants, where these types of failures were often solved by applying available heuristics. This contrasts with the higher frequency of *nature* attribution from unfamiliar tasks (see Figure 1).

4.3 Attribution behaviour

We now discuss how our participants searched for failure attributions. While any failure attributions could be selected (as attribution space) and tested, we observed a general behavioral pattern from their search. We classify this observed pattern into *internal* versus *external* attributions. We define an attribution as *internal* when a scientist considers the failure to be associated with solutions over which she has complete autonomy. A solution is under a scientist's autonomy to the extent that she does not manipulate or question the assumptions of the initial conditions of the experiment, such as the quality of a purchased sample, instrument, and even the validity of previous findings. On the other hand, *external* attribution considers solutions that question these initial assumptions. For example, a scientist questioning the quality of the material she made would be classified as *internal*. Meanwhile, attribution would be categorized as *external* when she questions the quality of the materials made by collaborators or manufacturers. Similarly, questioning the functioning of lab-built instrument systems would be *internal,* while questioning the instrument parts made by a manufacturer would be *external*. Also, questioning the procedures developed by a scientist would be *internal,* while questioning procedures developed elsewhere would be *external*.

Internal attribution: Is it my fault?

Figure 2 illustrates how often our participants started their search with internal versus external attributions. As seen from both subplots in Figure 2, most searches started with *internal* attributions. Examples include lab members casually blaming their "bad hands" after failing difficult procedures. Some participants would spend days or months trying to replicate published methods. Our participants rarely questioned the credibility of known procedures or purchased instruments immediately after failure. They often instead reflected on the range of problems and solutions over which they had control. Moreover, many of these troubles were not initially reported to their supervisor. This *internal attribution* would sometimes lead to wasteful search when *externaly attribution* would have been more appropriate. For example, one participant from Wet Lab was troubleshooting a recently purchased characterization instrument that was producing molecular weight distribution that resembled those coming from highly contaminated material. He spent over six months troubleshooting the instrument, trying various solutions within his control, then receiving help from the instrument maker over the phone, before deciding that the instrument needed to be sent back to the manufacturer.

External attribution: Is it the instrument? Is it the materials that I got from others? Could it be nature?

External attribution refers to failure attributions whose problems and solutions cannot be directly resolved by the problem-solver. The *external attribution* may question the validity of procedures, materials, and instruments that would otherwise be taken as valid under an *internal* search. Examples include questioning the purity of materials obtained from collaborators, doubting the reliability of off-the-shelf instruments, or suspecting the feasibility of production procedures from published works. We provide an example from one member of Wet Lab who struggled to replicate her former lab members' synthesis procedure.

"I first followed the procedure from a recent publication, but it never worked... I was able to go back to earlier publications. I ended up with an earlier lab member's dissertation. They do provide a "synthesis" chapter. This was from one of our former post-doc's dissertations. Replication didn't work out from the beginning. I worked with the post-doc and tried to follow her method. But I was getting lots of sizes (of the molecule). I could wash (separate) it to get it monodisperse. Eventually, I checked the guy who came before her, and saw his method. He did it differently. Then, no problem. I could always get the exact size."

Our participant first tried to replicate a former lab member's chemical synthesis. When it failed, she began to dig into the literature, searching for useful methods sections among cited publications. Even after reading a former lab member's dissertation, she could not produce the material with a desirable size distribution, until finally meeting another previous postdoc in person, who was able to demonstrate the synthesis procedure that "worked" for her research. This example illustrates how our participants often spent significant time experimenting with various parameters within their control before ultimately questioning the validity of the materials, procedures, or instruments. The *external attribution* is also where the *nature* failure category may be considered. For instance, in one case, repeated failure followed by considering *nature* led to the discovery of a new material property. This is also the path by which published findings may come to be questioned by the scientific community after repeated failures to replicate: in other words, after a series of internal attributions and problem solving searches do not succeed, the attribution may change to *nature*, in this case meaning that the published result is not in fact valid (Reich 2009).

Situated knowledge and attribution

Scientists often incorporated situated knowledge in the attribution process. For example, knowing the age of a particular bottle of a supply chemical could be an important clue as to whether *material quality* was or was not a reasonable attribution, as was knowledge of the supplier's quality control reputation. Similarly, knowing the maintenance and modification history of a specific piece of equipment could be an important clue to the *instrument* versus *instability* attribution (Kusterer 1978; Orr 1996). One example from the Dry Lab member's early detection of a flawed sample can exemplify the role of situated knowledge in the attribution search process.

"Meanwhile, she showed me a "control" PMMA, which had undergone all procedures except the infiltration with TMA (Trimethylaluminium). Showing how "hazy" her PMMA sample is, she explained that the haziness indicated inadequate drying of the sample, contrasting with the "yellowish" appearance of correctly infiltrated material. When asked about her diagnostic process, she answered: 'it took a while to figure it out, although I suspected that. I suspected either the PMMA wasn't dry enough or TMA was coming out in the middle of the process and meeting water. I tried both and figured it out eventually. The drying process that we were using was sufficiently dry for thin-film but not this material (around 0.5 mm thin).'"

In this example, she was able to identify that her PMMA (Polymethyl Methacrylate) sample was not correctly infiltrated by observing its hazy appearance, which was different from the expected "yellowish" color. This situated understanding of the material was a result of a close familiarity with the PMMA material system in her lab setting, developed over time through numerous observations and experiments, during which she was able to recognize that the standard drying process from her instrument, while adequate for thin films, was insufficient for thicker materials like her sample. Thus, the attribution process, and therefore the problem solving path, may depend heavily on situated knowledge, which would vary across lab settings.

The observed difference in the internal versus external attributions

Our observation indicates that most attribution searches begin with *internal attribution*. Out of 38 instances of failure observed during field visits, only 5 considered that the cause of the failure could be found by searching *externally* (Figure 2.A). Even after pulling every instance of failures, including those outside our field visits (n=82), we find that most of the search started with *internal attribution* (Figure 2.B). Thus, our evidence strongly indicates that scientists search for the causes of the failure *internally*. In the following section, we discuss potential explanations for this structuring of the search path.

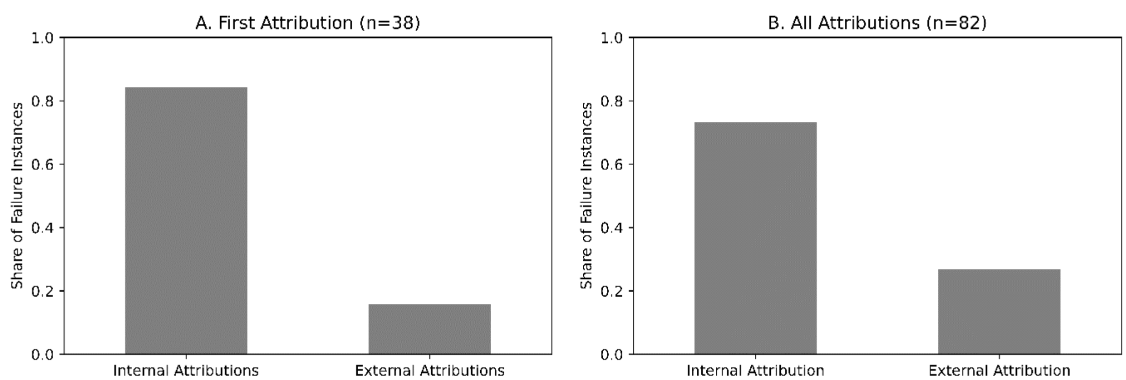


Figure 2. Internal versus external attributions. Note that subplot (A) corresponds to a relative frequency for each attribution behavior from the first failure entry, and subplot (B) corresponds to a relative frequency for each attribution behavior from the entire failure encounter. There were a total of 82 recorded data production failure entries. Of them, 38 instances were observed during our field visits.

Internalization of failure

Our evidence suggests that scientists tend to start their search within *internal attributions*, which raises questions about the internalization of failure in science. We provide several perspectives that could explain this tendency. One perspective is that the internalization may simply be reflecting scientists' reaction to strong priors about existing methods or findings. This Bayesian learning approach would suggest that prior beliefs influence the amount of evidence required to refute the existing findings (Dunbar 1995; Klahr 2002). Thus, unexpected outcomes from experiments would be perceived as inadequately executed experiments when the methods and findings are well-established.

It is also possible that the internalization of failure, specifically, accepting failures as part of their routine activities (Delamont and Atkinson 2001; Wylie 2019), maybe a cultural trait that scientists acquired through the enculturation process during their early training. According to this perspective, scientists are socialized to accept that real-world science is messy and that data production failure may be a default outcome. While this enculturation is partly intended to remedy the emotional distress associated with repeated failure during the training period, the consequence would be that scientists internalize data production failure. From this perspective, it also follows that reaching out for external help (such as reporting to their supervisor) before exhausting internal attributions would signal their incompetence. The internalization of failure could also reflect the enormous trust that scientists put into a system that governs the production of science through "organized skepticism"(Merton 1973). For example, in the face of data production failure, questioning the validity of published methods before carefully considering internal attributions may cause unwarranted reputation damage to the original authors and communities of scientists who have endorsed and relied on these methods for their research. While validating these perspectives is beyond the scope of this paper, what is clear from our observations is that scientists tend to internalize data production failure. The extent to which this behavior is unique to scientists is an open question. It also invites incorporation of attribution theory from social psychology literature, which examines the perceived causes of one's own behavior (Kelley and Michela 1980).

4.4 Selecting solutions

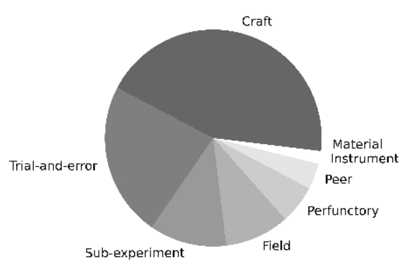
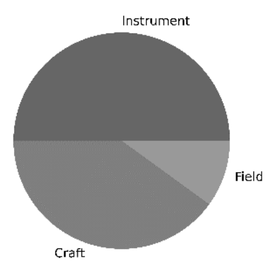
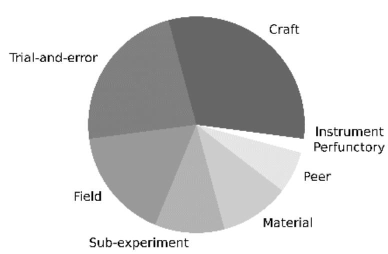
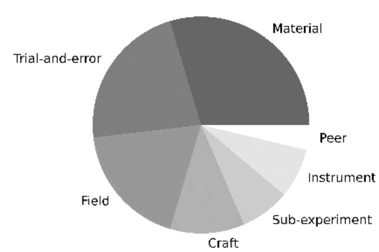
A range of solutions can be considered with respect to the attribution space constructed by our participants. We consider this set of solutions as a solution space. Table 1 reports eight different categories and descriptions based on attempted solutions observed from our field notes. These solutions are: craft, trial-and-error, field, sub-experiment, instrument, material, peer, perfunctory. Before describing these solutions in detail, we want to point out that our participants often applied solutions reflecting two distinct behaviors. Firstly, some failure attributions are familiar enough that they had read-to-apply solutions. They would apply solutions immediately without further investigating the failure. Inspired by problem-sovling literature (Newell and Simon 1972), we consider this behavior as a *heuristic approach*. Examples of a *heuristic approach* include ordering new parts for the failed instrument or purifying contaminated samples based on their available heuristics. Thus, these problems constitute cases where failures already have ready-made solutions from the perspective of our participants. When a *heuristic approach* is not available, our participants would generate additional information by constructing and iterating through sub-problems (Simon and Newell 1971). One reason for this iteration is that our participants would often construct multiple failure attributions associated with a single cause of the failure. Thus, narrowing down potential attributions would often require further information, which could be done by deliberately devising a *sub-experiment* or consulting their *peers* (see Table 1). We categorize this type of solution approach as *information-gathering*. Our observations suggest that the solutions from Table 1 could be be applied in either a *heuristic* or *information-gathering* approach, or both. It is important to note that more than one type of solution was often used to solve a single data production failure.

Table 1. Solution space

Solutions	Descriptions
Craft	*Craft* solutions involve technical and situated knowledge. They often facilitate identifying other solutions. Craft solutions can include domain-specific method (Simon, Langley and Bradshaw 1981) and situated knowledge (Orr 1996; Suchman 2006; Tyre and Von Hippel 1997).
Trial-and-error	*Trial-and-error* are unstructured or sometimes random attempts to modify or repeat procedures or change materials to solve the failure.
Field	*Field* involves the use of domain-specific knowledge involving a deep structural and scientific understanding behind data production.
Sub-experiment	*Sub-experiment* is an additional experiment used to generate further information to identify failure attributions.

Instrument	*Instrument* solutions include sending out instruments to manufacturers or ordering new parts.
Material	*Material* solution involves modifying or using new material.
Peer	*Peer* solutions involve asking for advice and feedback from colleagues or advisors. Their information could either expand or narrow down both attribution and solution space.
Perfunctory	*Perfunctory* solutions are a type of heuristics that are generally known to work without any logical explanation.

Table 2. Solutions by task familiarity and attributions

	Internal	External
Familiar		
Unfamiliar		

Familiarity, search, and solutions

Instead of providing a detailed description of every application of each solution, we highlight a selection of notable examples. Table 2 illustrates the distribution of solutions (see Table 1) by familiarity with tasks and attribution behaviors. *Craft* solutions, by far, were the most common and widely used solutions (see Table 2). *Craft* solutions involved our participants' technical and situated understanding of procedures, materials, instruments, and their complex interactions. Other than its self-explanatory role of making data production more efficient and reliable, *craft* knowledge helped our participants to make sense of and use information generated from applying other solutions. In other words, craft knowledge facilitated the use of *heuristics* apporaches. This was particularly evident when one participant from Dry Lab was troubleshooting the reactor she had built. Upon identifying gas leaks, her craft knowledge facilitated using the existing protocol to identify the source of the leak. This involved devising *sub-experiments* to determine whether the leak was caused by a faulty valve, gauge, or chamber. Another example from Wet Lab includes our participant's ability to quickly devise an artificial atmosphere around his instrument to limit the evaporation of theta solvent, which was limiting his data production.

Trial-and-error was another common solution used by our participants (see Table 2). Interestingly, this solution was more often sought when a participant's attribution space was depleted. For example, simply repeating a tricky chemical synthesis was a common *trial-and-error* solution from our participant who attributed the failure *internally* to familiar data production failure (as shown in the upper-left corner of Table 2). Note that this participant considered various potential attributions and solutions for the failure before resorting to *trial-and-error*. This behavior is consistent with the prior laboratory experiments, indicating that repetitive behavior is more common with uncertainties in data quality or when working hypotheses are depleted (Gorman 1986; Gorman 1989; Klahr and Dunbar 1988). While our participants agreed that repetition can improve *craft knowledge* (Delamont and Atkinson 2001; Peterson 2015), understanding the mechanisms behind "successful" data production was often challenging. Our field notes provide a clear illustration of the challenge of learning from unstable data production failures.

"For example, cobalt is extremely difficult to replicate. We have even reached out to these authors. It turns out they left steps out of the description, whether it is a probe v. a bath, that you have to

continuously bubble through nitrogen or argon. But, if it is too high, you get frothing. It took me almost one year to replicate this research. I don't know how I got it to work. What happened is that we changed this glassware because I broke one. And it worked after that. So, I am blaming this on this glassware manufacturer (laugh). The truth is it is kind of hard to explain how I got this right… I asked the post-doc to help. She came and watched me and did some troubleshooting, but still not there. Finally, I hit the sweet spot. Using fresh reactants. Got the argon just right… Yes, I have been able to do it a couple of times. But, I don't know for sure why it happened."

As discussed before, multiple solutions were often used to address data production failures. One example illustrates how *trial-and-error, craft,* and *field* solutions are used when searching for *external* attributions to *unfamiliar* data production failure (lower-right corner of Table 2). In this case, one member of Wet Lab searched for ideal material systems that could demonstrate the general applicability of his novel instrumental method, which used the speed of sound measurements to probe conformational changes in macromolecular solutions. Interestingly, both his scientific *field* knowledge and his situated *craft* knowledge of the instrument guided the search for the right material systems during a more than 6-month period of search, where he considered examining a range of macromolecules known in theory to exhibit sharp conformational changes (such as lysozyme) or utilizing macromolecules already familiar to the lab (such as polypeptides).

Lastly, we discuss the *perfunctory* solution, which was mostly observed when our participants considered internal attributions to familiar data production failures (upper-left corner in Table 2). These are the solutions that worked before, but the scientists cannot explain how or why. Common examples include rebooting instruments or computer programs. For instance, one participant from Wet Lab resolved an optical experiment issue by restarting the instrument, which, according to him, was "everyone's favorite trick." Another example from Dry Lab shows *perfunctory* solutions applied to the instrument not reporting frequency output, which was addressed by applying two methods: 1) pushing down the sample holder of the chamber slightly and 2) restarting the program if the first trick does not work. Interestingly, many *perfunctory* solutions were associated with failure attribution, *don't care*, as they often didn't bother to know the cause of this type of failure. One important attribute of the *don't car*e category is that it is also a situated label. Depending on the purpose of the experiment, knowing why a particular step fails or works may be constructed as critical for understanding or may be constructed as irrelevant to the ultimate goal of that researcher's study. For example, in the quote above about missing information in the published paper, it is possible that the information was left out because monodispersion that was so critical to the respondent's experiment was irrelevant to the paper author's study. This distinction likely drives when perfunctory solutions are considered adequate or unsatisfactory.

4.5 The Cycle of Doubt: A behavioral model of addressing data production failure

The Cycle of Doubt

Based on our observations and analyses, we present a behavioral model of how scientists address data production failures. We refer to this model as the "Cycle of Doubt", to reflect the iterative and uncertain nature of the failure-coping process. The key concepts are familiarity, attribution space, attribution behavior, and solution space. Our observations suggest that solutions are seldom "solved" by the first attempt except for rare cases where a single *heuristic* approach was sufficient to solve the failure. More often, addressing failure involved iterations of the troubleshooting process illustrated in our model shown in Figure 3.

The model starts with *failure detection*. Note that we are beginning with a failure identification. It is also an interesting question to study when a failure is detected (versus having the output pass through the process as "data"), or when the output is questioned and then labeled as a failure versus plausible data, but that is beyond the scope of this analysis. In our case, the data production failures were sufficiently obvious that there was little hesitation on the part of the scientist as to whether this was a failure or not. For example, in the case of an incinerated or vaporized sample that was therefore unable to be characterized, the "Is it a failure?" question was not raised. Hence, we begin with a failure being declared. Our findings suggest that failures are often identified by detecting sensory cues (color, smell, texture, etc.) or symbolic (figures, numbers, and charts) representations. The failures encountered by scientists can be either familiar or unfamiliar, depending on their prior experience with failure-generating tasks. A troubleshooting scientist then moves to the *attribution space*, where potential failure attributions would be considered. Choosing one of these failure attributions would be *internal* if a scientist perceives that the failure can be resolved within her autonomy. The choice would be *external* if a scientist perceives that the failure is attributed to reasons beyond her autonomy. Note there is no one-to-one correspondence between internal versus external attributions and the six attributions we categorized. For example, a scientist questioning the quality of her material would be *internal*, but this attribution would be *external* if the material was made elsewhere.

Once failure attributions are selected, a troubleshooting scientist will then apply appropriate solutions from the *solution space*. As discussed before, solutions could be selected with a *heuristic* approach or with an *information-gathering* approach, depending on the condition of data production failure. After applying solutions, a troubleshooting scientist will evaluate the outcome, possibly altering her failure representation. By this point, a troubleshooting scientist can move on when failures are resolved or by ignoring the failure. Otherwise, she would repeat the process.

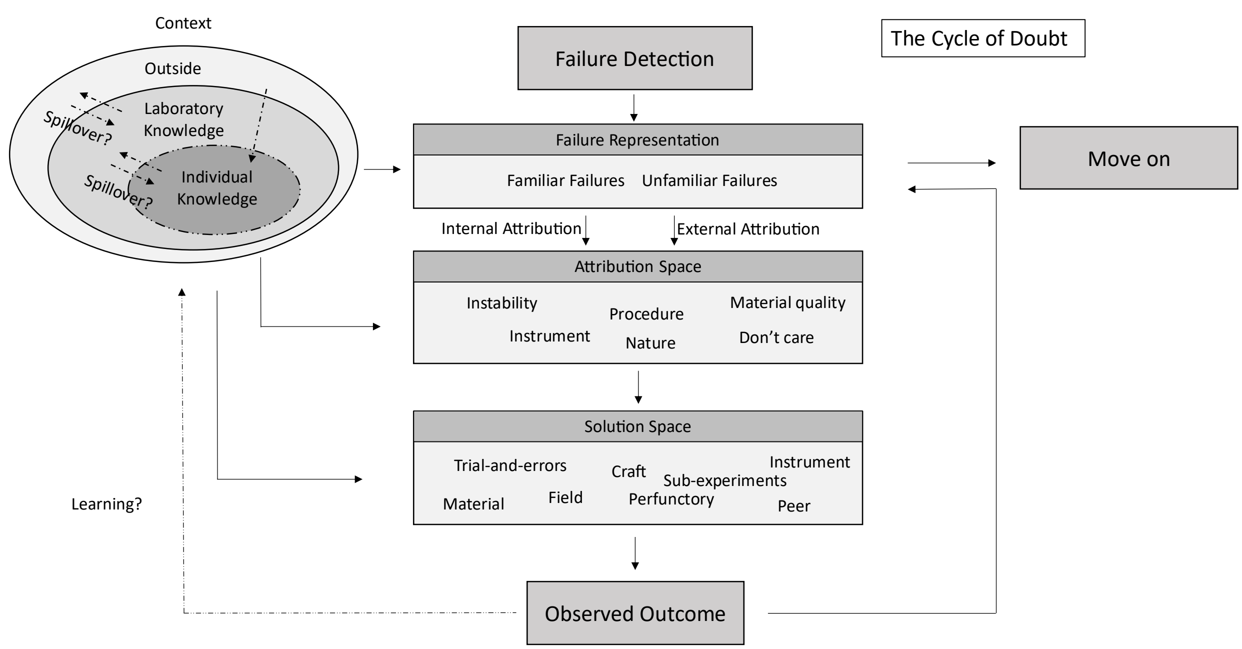


Figure 3. The Cycle of Doubt: A behavioral model of addressing data production failure

It is important to note that the Cycle of Doubt is shaped by the social and physical context in which data production failures occur. This context is shown in the upper-left corner of Figure 3. This region has different levels, including individual knowledge, laboratory knowledge, and outside knowledge. Arrows that connect this region to each stage of the Cycle of Doubt process represent the effects of social and physical context on the failure-solving process. Arrows that go back and forth between them indicate knowledge spillover – i.e., the construction and the shareability of the information generated within the lab. The learning from failure is illustrated by a dotted arrow that travels from the observed outcome to the context region to reflect often ambiguous feedback produced in the Cycle of Doubt process.

4.6 Conditions for successful failure-solving process

The cycle of doubt model emphasizes the iterative nature of reconstructing failure representations. However, our observations suggest that useful information is often scarce at each stage of the model due to two different types of uncertainties. Firstly, information could be scarce when exploring unfamiliar tasks, as making sense of the information produced would be difficult. Secondly, even familiar tasks often involve unstable data production processes, which make it difficult to rely on the information generated. Interestingly, successful failure-solving scientists tend to iterate the cycle faster because they can generate useful information faster. We discuss two interesting characteristics related to the ability to generate useful information.

One key aspect of successful failure-solving processes often involves quickly devising the right sub-experiments. This requires a combination of both craft and formal field knowledge around data production. Below, we provide an example of the application of a sub-experiment from Dry Lab to identify the source of the failure in a chemical reactor.

"To figure out it was leaking, I pumped it down and isolated the chamber… I measure this in an isolated state to see how much pressure is increasing. So, if it is increasing, there is probably a leak. But the problem with the leak is that it could be due to the atmospheric gas coming in, but it also could be water outgassing from the chamber wall. So, you have to wait to see if the increase is like a slope or like a plateau. The slope is a leak, and the plateau is water. If it is water, you can heat and wait. So, before I run the experiment every time, I check that. And a lot of times, I have to wait a while to get the water out. [She had several troubleshooting programs on her computer]"

Because of the frequent leaks from the chamber, she developed several customized programs to identify the source of the problem. These programs are situated around her specific instrument and her tasks, so they may be irrelevant to other instruments that have the same function. The example also shows how her scientific knowledge is used to identify different patterns of pressure output data. The programs she used and her ability to quickly identify the pattern became available heuristics, which was possible because she had both situated and field knowledge of data production. However, this type of heuristic was not always available among data production tasks that were highly unstable, such as those performed in an "open system." Because systematic ways to generate codified data were physically challenging, learning from failures was extremely difficult in such settings. Thus, the ease of solving data production failures may be highly dependent on the task environment.

The other aspect of successful failure-coping processes is the ability to make use of information generated from the lab. We found that labs with high overlap in data production tasks enjoyed high-spillover of useful information. For example, members of Dry Lab would often share lab-built chemical reactors. This overlap allowed some members of Dry Lab to use the information generated by other lab members to readjust their attribution space without having to run their own sub-experiments. The shared situated knowledge about specific instruments allowed some participants to rapidly produce and share useful information to quickly solve data production failures. Therefore, we can conclude that the conditions for a successful failure-solving process require the ability to generate useful information. This ability can be influenced by the level of situated knowledge, the task environment and the lab organization of data production work.

5. Discussions and Conclusions

In this paper, we presented ethnographic observations of an overlooked but critical dimension of the work of laboratory scientists – the frequent failure to produce usable data from their experiments. Our paper suggests that data production failure is common and that scientists consider a range of attributions and solutions following a fairly structured pattern. Specifically, we find that scientists tend to internalize failures by attributing the failure to themselves before attributing the failure to those outside of their control. While understanding the source of internalization of failure in science is beyond the scope of this paper, we find this to be an interesting characteristic as it may be a critical site to understand how data production can be accelerated.

Our paper presents mixed implications for the ongoing debate around the potential of AI and automation in science. On the one hand, we anticipate that the advancements in deep learning in the computational sciences (Ramprasad et al. 2017; Tshitoyan et al. 2019) can help scientists narrow down large combinatorial spaces to avoid dead-end searches, as a significant amount of time is often lost in finding the appropriate material systems. However, our findings suggest that the bulk of data production failures occurred due to instability in the production process. Even data production under a relatively well-controlled closed system often required our participants to sit around the instrument for hours to monitor frequent failures. One way to free scientists from "babysitting" instruments is to apply various analytics used in industrial sectors, such as installing general-purpose sensors that would collect information such as vibration, sounds, and temperature, thereby learning when failures would happen. As such, monitoring failure preemptively could potentially predict and reduce data production failures. Moreover, the explainability of algorithms used in analytics may depend on the purpose of data production tasks. If data production is the focus of the research, an explainable algorithm may be desirable, as opposed to if it is merely an intermediate step. However, due to similar problems posed by data production work done in open-system, we expect that the benefits from automation and the uses of data analytics would vary by subfield.

Our research also highlights the significance of the organization of data production work. Useful information in solving data production failures was localized and situated such that only those who had proximate backgrounds were able to make use of information spillover. Interestingly, this finding contrasts with the other prominent view that suggests how a diversity of members' expertise may be beneficial for problem-solving (Argote, Lee and Park 2020; Cummings 2004; Dunbar 1995) due to heterogeneous failure-attributions and potential solutions generated by a diverse background of expertise. Instead, our finding aligns with research that emphasizes the significance of a closed network with strong ties in work organization for transferring tacit knowledge (Hansen 1999).

Meanwhile, how situated knowledge and formal scientific knowledge are put into action to address frequent technical problems in our paper provides an implication for the organization of work in science. Interestingly, the labs we observed were solely staffed by graduate students and post-docs whose works also involved types of troubleshooting documented by previous studies of lab technicians (Barley and Bechky 1994; Bruyninckx 2020; Doing 2004). Given that all members of the labs utilized both situated and formal scientific knowledge to address failures, we are unsure whether this reflects the inseparability of skills for solving problems or the lack of division of labor for some other reasons. Moreover, while supervisors provided essential field expertise, many failures were resolved without supervisors, and, furthermore, supervisors often had no solutions for the cases brought to them, but instead suggested paths along the Cycle of Doubt (in essence, counseling, keep looking). Given that substantial graduate training involves mastering craft knowledge (Delamont and Atkinson 2001; Peterson 2015), it is not clear whether mastering such skill sets is a necessary condition for becoming an independent investigator.

Finally, we find that generating useful information is critical for addressing data production failure by accelerating the Cycle of Doubt process. In this sense, having a situated understanding around data production was important as field knowledge was hardly used in a vacuum to address data production failure. While task repetition was essential for developing a situated understanding of the data production, we find some data production tasks yielded insufficient information feedback to apply available heuristics or the ability to apply sub-experiments. This was particularly the case for much of the chemical synthesis conducted in open systems. Learning from failures is difficult in such conditions where some failures are often resolved by arbitrary variables such as replacing equipment or reagents. In contrast, data production failures from closed systems, such as chemical reactors, were much more tractable as our participants were able to systematically generate sub-experiments in relatively well-isolated conditions. Given such an apparent difference in instability in the data production tasks, we expect to observe substantial field and subfield variations in managing data production failures. Overall, these results show the prevalence of data production failures, the importance of situated knowledge for addressing such failures, and the contextual contingencies that affect access to such situated knowledge. These findings provide important guidance for understanding the work of science.

REFERENCES

Argote, Linda, and Eric Darr. 2000. "2 Repositories of Knowledge in Franchise Organ-izations: Individual, Structural, and." *The nature and dynamics of organizational capabilities* 2:51.

Argote, Linda, Sunkee Lee, and Jisoo Park. 2020. "Organizational Learning Processes and Outcomes: Major Findings and Future Research Directions." *Management Science*.

Barley, Stephen R, and Beth A Bechky. 1994. "In the backrooms of science: The work of technicians in science labs." *Work and occupations* 21(1):85-126.

Barley, Stephen R, and Gideon Kunda. 2001. "Bringing work back in." *Organization science* 12(1):76-95.

Barley, Stephen R. 1988. "The Social Construction of a Machine: Ritual, Superstition, Magical Thinking and other Pragmatic Responses to Running a CT Scanner." Pp. 497-539 in *Biomedicine Examined*, edited by Margaret Lock and Deborah Gordon. Dordrecht: Springer Netherlands.

Becker, Howard Saul. 2002. *Boys in white: Student culture in medical school*: Transaction publishers.

Bruyninckx, Joeri. 2020. "Somatic Vigilance and Sonic Skills in Experimental Plasma Physics." *Science as Culture* 29(3):450-73.

Burnell, Jocelyn Bell. 2004. "So few pulsars, so few females." *Science* 304(5670):489-90.

Cambrosio, Alberto, and Peter Keating. 1988. "“Going monoclonal”: Art, science, and magic in the day-to-day use of hybridoma technology." *Social Problems* 35(3):244-60.

Campbell, Robert A. 2003. "Preparing the next generation of scientists: The social process of managing students." *Social Studies of Science* 33(6):897-927.

Collins, Harry. 2010. *Tacit and explicit knowledge*: University of Chicago Press.

Cummings, Jonathon N. 2004. "Work groups, structural diversity, and knowledge sharing in a global organization." *Management Science* 50(3):352-64.

Cyert, Richard M, and James G March. 1963. "A behavioral theory of the firm." *Englewood Cliffs, NJ* 2(4):169-87.

Delamont, Sara, and Pauln Atkinson. 2001. "Doctoring uncertainty: Mastering craft knowledge." *Social Studies of Science* 31(1):87-107.

Doing, Park. 2004. "‘Lab Hands’ and the ‘Scarlet O’ Epistemic Politics and (Scientific) Labor." *Social Studies of Science* 34(3):299-323.

Dunbar, Kevin. 1995. "How scientists really reason: Scientific reasoning in real-world laboratories." *The nature of insight* 18:365-95.

Fujimura, Joan H. 1987. "Constructingdo-Able'problems in Cancer research: Articulating alignment." *Social Studies of Science* 17(2):257-93.

—. 1988. "The molecular biological bandwagon in cancer research: Where social worlds meet." *Social Problems* 35(3):261-83.

—. 1992. "Crafting science: Standardized packages, boundary objects, and “translation.”." *Science as practice and culture* 168(1992):168-69.

Gorman, Michael E. 1986. "How the possibility of error affects falsification on a task that models scientific problem solving." *British Journal of Psychology* 77(1):85-96.

—. 1989. "Error, falsification and scientific inference: An experimental investigation." *The Quarterly Journal of Experimental Psychology Section A* 41(2):385-412.

Hackett, Edward J. 1990. "Science as a vocation in the 1990s: The changing organizational culture of academic science." *The journal of higher education* 61(3):241-79.
Hagstrom, Warren O. 1964. "Traditional and modern forms of scientific teamwork." *Administrative science quarterly*:241-63.
Hansen, Morten T. 1999. "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits." *Administrative science quarterly* 44(1):82-111.
Harper, Douglas. 1987. *Working knowledge: Skill and community in a small shop*: University of Chicago Press.
Jordan, Kathleen, and Michael Lynch. 1992. "3. The Sociology of a Genetic Engineering Technique: Ritual and Rationality in the Performance of the" Plasmid Prep"." Pp. 77-114 in *The right tools for the job*: Princeton University Press.
Kelley, Harold H, and John L Michela. 1980. "Attribution theory and research." *Annual review of psychology* 31(1):457-501.
Klahr, David. 2002. *Exploring science: The cognition and development of discovery processes*: MIT press.
Klahr, David, and Kevin Dunbar. 1988. "Dual space search during scientific reasoning." *Cognitive science* 12(1):1-48.
Klahr, David, and Herbert A Simon. 1999. "Studies of scientific discovery: Complementary approaches and convergent findings." *Psychological Bulletin* 125(5):524.
Knorr-Cetina, Karin. 1981. "The Manufacture of Knowledge."
—. 2009. *Epistemic cultures: How the sciences make knowledge*: Harvard University Press.
Kuhn, Thomas S. 1962. *The structure of scientific revolutions*: University of Chicago press.
Kusterer, Ken C. 1978. *Know-how On The Job: The Important Working Knowledge Of"" Unskilled"" Workers*: Westview Press.
Latour, Bruno. 1987. *Science in action: How to follow scientists and engineers through society*: Harvard university press.
Latour, Bruno, and Steve Woolgar. 1979. *Laboratory life: The construction of scientific facts*: Princeton University Press.
Lave, Jean, and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*: Cambridge university press.
Lynch, Michael. 1985. *Art and artifact in laboratory science: A study of shop work and shop talk in a research laboratory*: Routledge.
Merton, Robert K. 1957. "Priorities in scientific discovery: a chapter in the sociology of science." *American Sociological Review* 22(6):635-59.
—. 1973. "The normative structure of science." Pp. 267-78 in *The sociology of science: Theoretical and empirical investigations*, edited by Norman W Storer: The University of Chicago Press.
Nelson, Richard R, and Sidney G Winter. 1981. *An Evolutionary Theory of Economic Change* harvard university press.
Newell, Allen. 1967. "Studies in problem solving; subject 3 on the crypt-arithmetic task Donald+ Gerald= Robert."
Newell, Allen, and Herbert Alexander Simon. 1972. *Human problem solving*: Prentice-Hall Englewood Cliffs, NJ.
Orr, Julian E. 1996. *Talking about machines: An ethnography of a modern job*: Cornell University Press.
—. 1998. "Images of work." *Science, technology, & human values* 23(4):439-55.

Owen-Smith, Jason. 2001. "Managing laboratory work through skepticism: Processes of evaluation and control." *American Sociological Review*:427-52.
Peterson, David. 2015. "All that is solid: Bench-building at the frontiers of two experimental sciences." *American Sociological Review* 80(6):1201-25.
Ramakrishnan, Venki. 2018. *Gene Machine: the race to decipher the secrets of the ribosome*: Basic Books.
Ramprasad, Rampi, Rohit Batra, Ghanshyam Pilania, Arun Mannodi-Kanakkithodi, and Chiho Kim. 2017. "Machine learning in materials informatics: recent applications and prospects." *npj Computational Materials* 3(1):54.
Reich, Eugenie Samuel. 2009. *Plastic fantastic: How the biggest fraud in physics shook the scientific world*: Macmillan.
Ribeiro, Barbara, Robert Meckin, Andrew Balmer, and Philip Shapira. 2023. "The digitalisation paradox of everyday scientific labour: How mundane knowledge work is amplified and diversified in the biosciences." *Research Policy* 52(1):104607.
Shapin, Steven, and Simon Schaffer. 1985. *Leviathan and the air-pump: Hobbes, Boyle, and the experimental life (New in paper)*: Princeton University Press.
Simon, Herbert A. 1947. *Administrative behavior*: Simon and Schuster.
—. 1992. "Scientific discovery as problem solving." *International Studies in the Philosophy of Science* 6(1):3-14.
Simon, Herbert A, Patrick W Langley, and Gary L Bradshaw. 1981. "Scientific discovery as problem solving." *Synthese*:1-27.
Simon, Herbert A, and Allen Newell. 1971. "Human problem solving: The state of the theory in 1970." *American Psychologist* 26(2):145.
Star, Susan Leigh. 1983. "Simplification in scientific work: An example from neuroscience research." *Social Studies of Science* 13(2):205-28.
Suchman, Lucy. 2006. *Human-Machine Reconfigurations: Plans and Situated Actions*: Cambridge University Press.
Tshitoyan, Vahe, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A Persson, Gerbrand Ceder, and Anubhav Jain. 2019. "Unsupervised word embeddings capture latent knowledge from materials science literature." *Nature* 571(7763):95-98.
Tyre, Marcie J., and Eric Von Hippel. 1997. "The Situated Nature of Adaptive Learning in Organizations." *Organization science* 8(1):71-83.
Von Hippel, Eric. 1994. "“Sticky information” and the locus of problem solving: implications for innovation." *Management Science* 40(4):429-39.
von Hippel, Eric, and Marcie J. Tyre. 1995. "How learning by doing is done: problem identification in novel process equipment." *Research Policy* 24(1):1-12.
Whalley, Peter, and Stephen R. Barley. 1997. "TECHNICAL WORK IN THE DIVISION OF LABOR

STALKING THE WILY ANOMALY." Pp. 23-52 in *Between Craft and Science*, edited by Stephen R. Barley and Julian E. Orr: Cornell University Press.
Wylie, Caitlin Donahue. 2019. "Socialization through stories of disaster in engineering laboratories." *Social Studies of Science* 49(6):817-38.
Zuboff, Shoshana. 1988. "In the age of the smart machine."